ROLES OF VISUAL WORKING MEMORY, GLOBAL PERCEPTION AND EYE-MOVEMENT IN VISUAL COMPLEX PROBLEM SOLVING

by

Xiaohui Kong

B.S. in Computer Science, Zhejiang University, China P.R., 2002

Submitted to the Graduate Faculty of

Arts and Science in partial fulfillment

of the requirements for the degree of

PhD in Intelligent Systems

University of Pittsburgh

2009

UNIVERSITY OF PITTSBURGH

SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Xiaohui Kong

It was defended on

August 4th 2009

and approved by

Prof. Stephen Hirtle, Professor, School of Information Sciences

Prof. Michael Lewis, Professor, School of Information Sciences

Prof. Paul Munro, Associate Professor, School of Information Sciences

Prof. Garrick Wallstrom, Assistant Professor, Department of Biomedical Informatics

Dissertation Advisor: Prof. Christian Schunn, Associate Professor, Department of Psychology, Intelligent Systems Program

ROLES OF VISUAL WORKING MEMORY, GLOBAL PERCEPTION AND EYE-MOVEMENT IN VISUAL COMPLEX PROBLEM SOLVING

Xiaohui Kong, PhD

University of Pittsburgh, 2009

In this dissertation, I explore roles of visual working memory, global perception and eye-movement in complex visual problem solving. Four experiments were conducted and two models were built and tested. Experiment one and model one showed that global information plays an important role and there is an interaction between external representation and internal VWM on global information representation. Experiment two and model two showed that this interaction is achieved by encoding global information with eye-movements throughout the duration of solving a problem. A very regular eye-movement pattern is observed in experiment two. Experiment three further tested the hypothesis that this eye-movement pattern is a result of the individual's VWM limitation by measuring the correlation between individual differences in the quantitative features of the eye-movement pattern and VWM size. The second model assumes that global and local information share a unified VWM capacity limitation. In the fourth experiment, I tested this hypothesis along with several alternative hypotheses. Results of the fourth experiment support the unified capacity hypothesis best and thus make a complete story for the interaction between VWM, global information processing and eye-movements in complex visual problem solving.

Even with such a limited amount of VWM capacity, human visual cognition is able to solve complex visual problems by keeping a balanced amount of global and local information in

VWM. This balance is achieved by eye-movements that encode both types of information into a unified VWM. Thus, although VWM has such a limited capacity, through frequent eye-movements, visual cognition is able to encode complex visual information in a temporal manner. At each instance, the amount of information encoded is limited by the capacity limitation of VWM but the global information encoded can further guide eye-movements to acquire information that is needed to make the next decision.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

The human visual system is one of the most complex information processors. It supports a vast majority of human intelligent behaviors. Significant amounts of research had been performed to investigate human vision for over a century in a variety of disciplines including cognitive psychology, neuroscience, artificial intelligent, cognitive science and etc. A lot of interesting concepts related to the visual system have been discovered, such as ensemble visual information [1] and etc. Yet, many underlying mechanisms of this complex system remain unknown. For example, one open puzzle involves the fundamental limitations of the visual system. A vast amount of information reaches the retina at each moment, yet researchers have found that only a minor fraction of it is encoded and stored temporally. The concept of visual working memory was first discussed along with the concept of general working memory [2]. In the most recent decade, it was argued that there are only several chunks of information can be manipulated at the same time in human cognition [3-8].

Although only a very limited amount of visual information is represented at any certain instance, researchers have shown that humans are able to solve complex visual problems in a short time. This phenomenon is intuitively consistent with our daily experience in utilizing our visual problem solving ability. A variety of heuristic algorithms [9-12] were proposed to explain this phenomenon from a computational time complexity point of view.

Not only can human visual system solve complex visual problems in a very short time but also it solves them under VWM constraint. How can such a limited VWM capacity support

such complex visual problem solving processes? That is, how could the human visual system be so powerful and yet so limited at the same time? The curiosity to explore possible answers for these questions sparked this dissertation research.

The outline of this dissertation goes as following. In the next chapter, chapter 2, I will review some background literature on visual working memory and human complex problem solving. Then in chapter 3, our first experiment and its results are discussed centering on the role of global information in complex visual problem solving. Based on these results, our first TSP model was built and discussed in this chapter as well. In chapter 4, two eye-tracking experiments and their results were discussed. I will discuss the relationships between visual working memory and global perception in the context of complex problem solving. In particular, I will describe our second model of human TSP solving that integrates both concepts. Those results further lead to chapter 5 in which experiment four and its results were described and discussed. Experiment four further tested the hypothesis that global and local information share a unified VWM capacity along with several alternative hypotheses regarding how global and local information are represented in VWM.

2.0 BACKGROUND

2.1 MEMORY: SHORT AND LONG TERM

Memory, as the means to carry information from one temporal state to another, is a crucial part of our cognition that supports most if not all intelligent behaviors. Memory makes what happens transiently in the outside world longer lasting internally for use by cognition. It is one of the central topics in cognitive psychology. There is a vast literature under this broad topic.

There are many ways to divide memory into sub-systems. For example, memory is divided, according to input sources or forms of representation, into visual memory, verbal (conceptual) memory, olfactory memory, etc. Besides the different experiment paradigms that are used to study these different memory subsystems, there is evidence supporting the separation of these subsystems in human cognition. For example, brain damage may result in loss of verbal or visual memory but not the other [13].

Another way to categorize memory is according to its temporal duration, dividing it into short-term memory and long-term memory. This distinction has been made since the 19th century [14, 15]. But it is only in recent decades, with the help of computerized experiment paradigms, that researchers have made significant advances in studying visual short-term memory, a particular memory system in our cognition that temporally stores visual information.

2.2 VISUAL WORKING MEMORY

The terminology “working memory” comes from two separate sources according to Baddeley [16]. One originated from the source of short term memory as a capacity limited storage in the human memory system that temporarily holds information. The other comes from an early classic computational model of intelligence [17], where “working memory” holds the production system. The concept of working memory in cognitive science nowadays inherit mainly from the concept of a limited storage in short term memory but more focused on its function of manipulation and computational storage as used in the production system. The term “visual working memory” and “visual short-term memory” are often used interchangeably with the label choice often depending on the experiment context. For a review of visual memory, please refer to [8].

Different from visual long-term memory, visual working memory is short-term, holding visual information for only several seconds unless refreshed through additional attention, whereas visual long-term memory may persist for an indefinite amount of time [8]. Encoding information into visual working memory is at a much faster rate, 20-50 ms/item, than into visual long-term memory [8, 18-20]. Researchers have also shown that visual working memory is highly limited in term of capacity [2, 21, 22]. One most widely accepted theory suggests that the capacity of visual working memory is limited to 2-7 fixed slots [4, 5, 7, 23]. Two major paradigms of experiment are usually used to study visual working memory (described in the next two sections). Evidence from both of these paradigms is summarized below.

2.2.1 The change detection experiment paradigm for studying visual working memory

The first experiment paradigm is based on a change-detection task. Many variation of this task had been used to study different aspects of visual working memory, but the main paradigm involves the following steps:

First, a sample array of objects with different features is briefly presented on the screen for the observers to try to remember. These objects are usually basic geometrical shapes and last on the screen for about 100ms to 500ms. Each object in the array may differ from each other in one or more features, such as shape, orientation or color. During this step, observers encode as many objects into visual working memory as possible.

Then after about 1s of a blank screen retention period, an array of objects is displayed on the screen for the second time. This time, there is either a change in the array, such as color change of an object, or the array is identical to the one displayed in the first step. The task of the observer is to tell whether a change has occurred. In most experiments, true and false trials occur in a 50/50 ratio and are randomly intermixed.

If an object is encoded in visual working memory, then presumably, an observer should be able to correctly detect a change in it. If an object was not encoded, then an observer has to rely on blind guessing, which has a 50% chance of being correct. So if an observer has a visual working memory capacity capable of holding K objects and N <= K objects are displayed, then the observer should be able to perfectly detect any change. Or in the case of N greater than K, then either the changed object is in one of the K objects in visual working memory with probability of K/N, or the object is not encoded with probability 1-K/N. So if an observer has a visual working memory capacity K, then in about K/N of all trials with size N the observer

should response with all correct answers. In the rest 1-K/N trials with size N, the observer should response correctly on half of these trials. This experiment paradigm and calculation of visual working memory capacity was first formalized by Pashler [24]and improved by Cowan et al. [6]. The resulting formula, which is often referred as Pashler-Cowan K formula, has an elegant simple form: K = N (H - F), where H is the hit rate and F is the false alarm rate.

In a classic experiment by Luck and Vogel's [3], a sample array consists of one, two, three, four, eight, or 12 color blocks. Blocks only differ from each other by their color. In 50% of trials, one block's color changed in the test array. When the number of blocks is below 3, the percentage of correct answers was almost 100. When the number of blocks is above 3, percentage of correctness started to decrease systematically. Using Pashler-Cowan K formula, visual working memory capacity is calculated to be around 3, which is consistent with the observation.

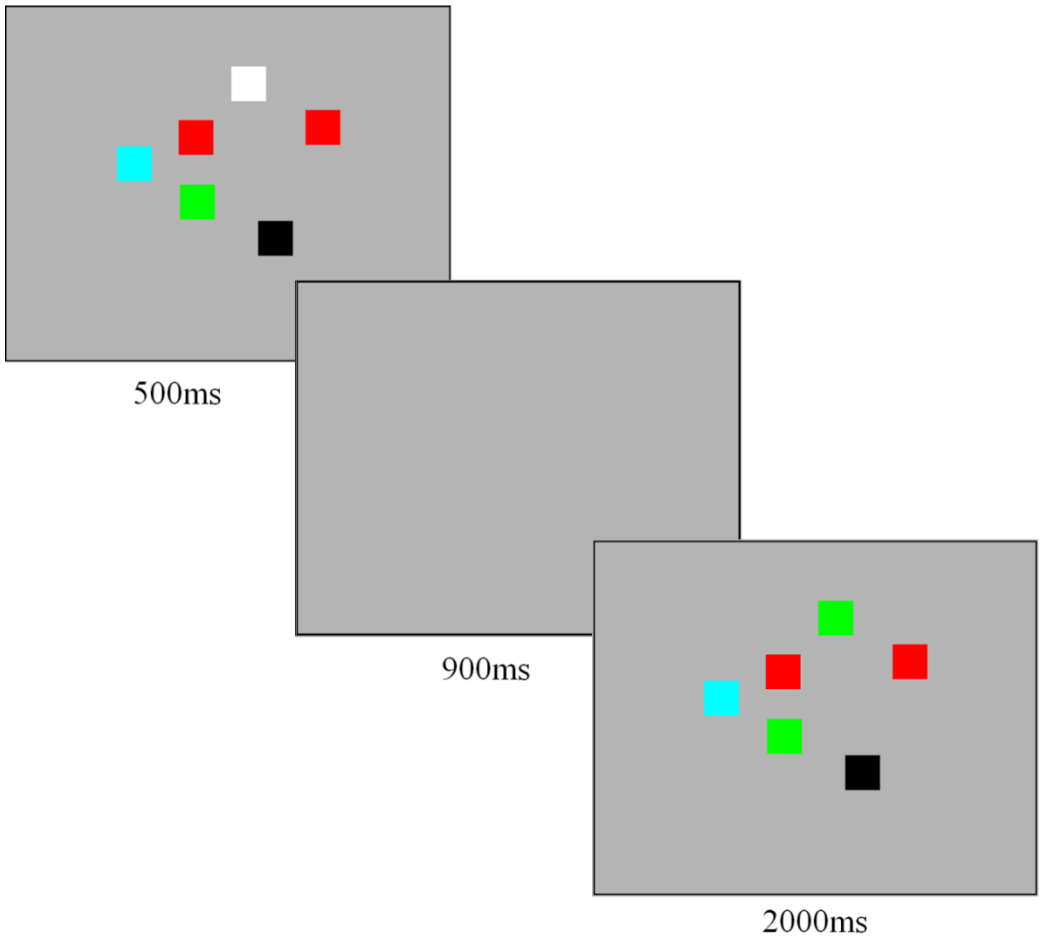


Figure 2-1: Example of a change-detection trial. Memory array stayed on screen for 500ms. After a 900ms retention interval, a test array appears on screen for 2s, waiting for a user response.

In visual working memory, one essential element of information is the spatial location of objects. During a search task, for example, visual working memory allows the mind to keep track of locations that have been previously attended. One phenomenon, often referred as inhibition-of-return, is that shifting attention to a just attended spatial location is slower than to a novel one [25, 26]. One explanation for inhibition-of-return is that attention is biased towards locations not encoded in visual working memory in visual search tasks. Data in several studies support this view [27-29]. For example, inhibition was found to occur to areas attended in the last few

seconds [27, 28]. If visual working memory is occupied by a concurrent task, the inhibition is reduced [29].

Despite the large amount of research in visual working memory, its role and interaction with other parts of cognition are not well understood. Presumably, visual working memory temporally holds some highly abstracted visual information so that other part of cognition can access them and operate on them. For example, possible operations include comparison, rotation, etc.

A visual puzzle demonstrating the limited capacity of visual working memory is shown in Figure 2-2. There are 16 umbrellas in this puzzle. The goal is find a pair of umbrella identical to each other. This identical pair of umbrella can be in a rotated position from each other but not mirrored. Solving this kind of puzzle takes quite a long time, since one can only stores a very limited amount of visual information in visual working memory at each moment. In this case, presumably, the color blocks making up the umbrellas are stored as chunks of information in visual working memory.

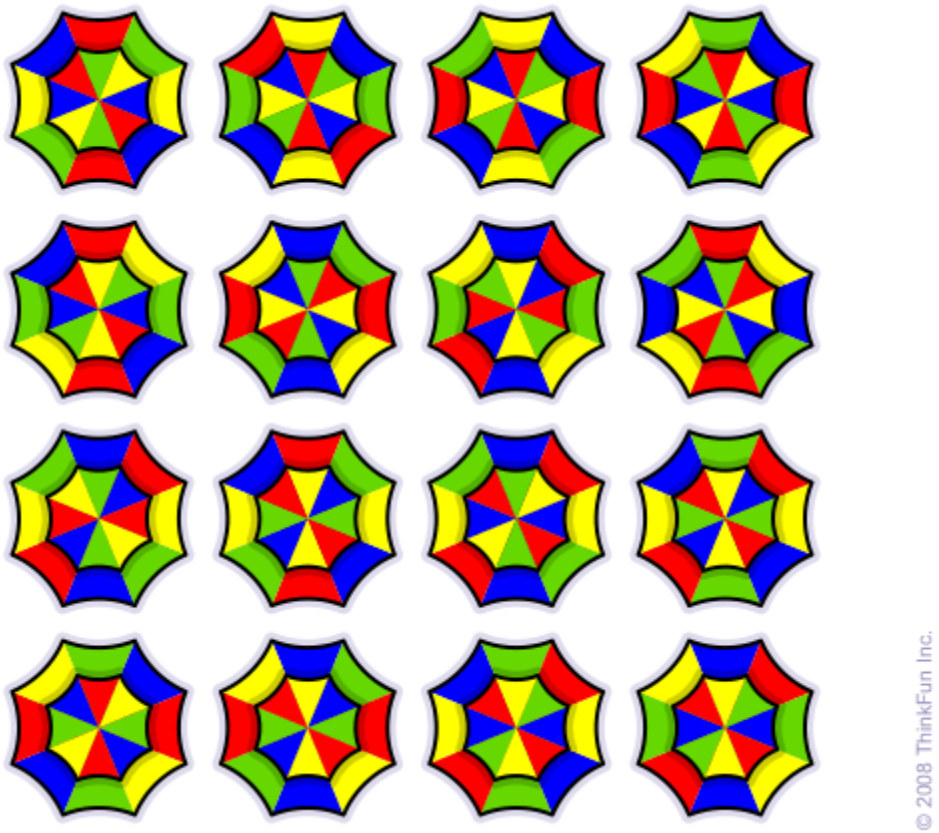


Figure 2-2. Visual puzzle: Find the pair of identical umbrellas

Measuring the capacity limitation of visual working memory across different dimensions has been a focus of research for the last two decades and has born a number of fruitful results [3, 4, 30-33]. The current most dominant theory of visual working memory capacity posits that visual working memory is measured in term of fixed number of "slots" and its capacity varies across individuals from 2 to 7 slots [3, 4, 31, 33, 34]. This capacity limitation is consistent with studies using the Multiple Object Tracking paradigm, described in the next section.

2.2.2 The multiple-objects-tracking paradigm and the FINST theory

The multiple-objects-tracking paradigm was originally developed by Pylyshyn along with the FINST model of attention [35, 36]. In a trial of a multiple-objects-tracking experiment, a certain number of simple objects, such as dots, move around on a computer screen. Within those objects, some are targets and the rest are distracters. The type of object is identified by a salient feature, such as color, in the beginning of the trial. Several tasks can be used to evaluate how well an observer is able to keep track of the objects. For example, one experiment may ask the observer to classify an object as a target or distracter when these types of objects became visually identical and moved around for a while (see Figure 2-3).

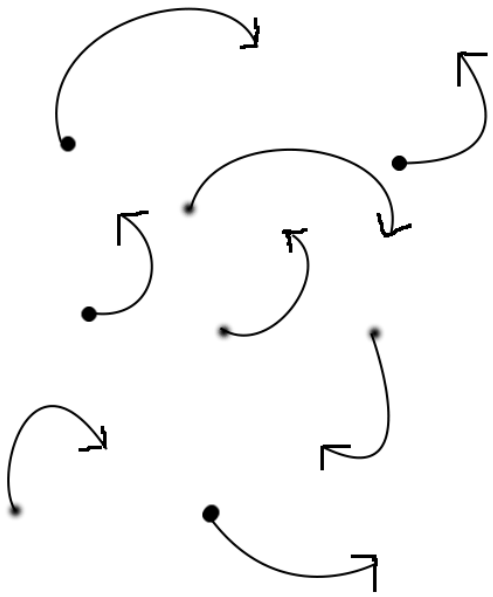


Figure 2-3: An example of multiple-object-tracking trial.

Studies have shown that people are able to keep track of about 4-5 objects simultaneously [35-38]. This number is consistent with the visual working memory capacity measured in the change-detection paradigm [4]. Although it is still controversial whether the multiple-objects-tracking task and the change-detection task share the exact same attentional resources [39, 40], the observation that both tasks utilize the same resources to temporally hold spatial locations is supported by many studies [41-45].

In real world visual problem solving, the ability to track moving objects and temporally hold object locations is critically important. Another very important visual ability is to process

global or statistical information. Human ability to process global visual information was described by Gestalt psychologist almost a century ago [46]. But it is only in recent years that global or statistical visual information processing has been formally studied within the framework of attention and visual working memory.

For example, Alvarez and Oliva [1] studied statistical visual information processing using a multiple-objects-tracking task. Results of three experiments were reported. All three experiments employed a dual-task paradigm combining multiple-object-tracking and missing item localization, although with variations across experiments. In the experiments, eight dots moves around the screen for a short duration (between 6s and 10s). Four of the dots are targets and four of them are distracters. The primary task (multiple object tracking) is to keep track of four moving target dots on the screen and count the total the number of times the dots touch two diagonal lines displayed on the screen (Figure 2-4). After all the dots stop moving, there is a 200ms interval with a blank display on the screen. Then all the points but one missing item reappear on the screen. The secondary task (missing item localization) is to locate where the missing item was before it disappeared. The missing item could be either a dot or a group of four dots (targets or distracters). If a group of four dots are missing, then the task is to locate where the centroid of this group of dots was. On the other hand, if one point is missing, then the task is to locate where the missing dot was.

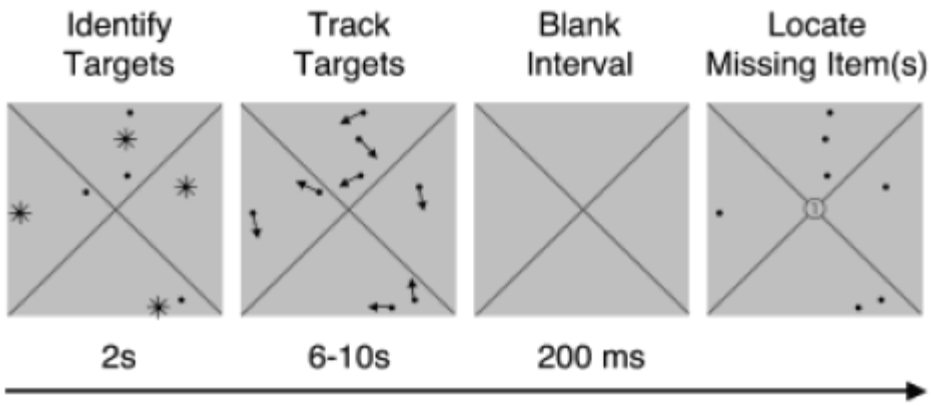


Figure 2-4. Multiple-object-tracking and missing-item-localization dual task paradigm. (Adapted from [1])

Results of three experiments were reported by Alvarez and Oliva [1] under this paradigm. In the first experiment, targets and distracters share the same color and the same spatial region in which they move around. In the second experiment, targets and distracters differed by color but share the same spatial region. In the third experiment, targets and distracters moved in different spatial regions but share the same color. Those three experiments yielded very similar results. In those experiments, participants could locate the centroid position of the distracter group at rates better than chance, but they couldn't locate a distracter point above chance level (except that in their first experiment, participants were also able to locate a position of an individual distracter better than chance). The ability to locate the centroid position but not the individual points suggests that it is possible for the visual system to encode global information without encoding every local element in it. Presumably, this ability is critically important for visual complex problem solving under a very limited visual working memory constraint.

2.3 HUMAN VISUAL COMPLEX PROBLEM SOLVING

Despite an extremely limited visual working memory capacity, the human visual system is a very complex and powerful information processor. Many tasks of visual problem solving that are trivial for human vision posted the hardest problems for artificial intelligence. However, many problems which are trivial for human vision are hard for computers because they are either ill-defined or need intensive interaction with other parts of human intelligence. For example, humans can easily navigate through a terrain with many obstacles but robots have a hard time to do it. Though the navigation task heavily depends upon visual perception, it also depends on peripheral sensors feedbacks and motor programming. Recognizing objects or people in a natural scene is another trivial task for human vision that is hard for computer vision.

From one direction, computer scientists have been working hard to build computer vision applications for different tasks. From the other direction, cognitive psychologists and neuroscientists have been studying the fundamental mechanisms underlying the human visual system. When studying the fundamental mechanism of human visual system, most tasks used in laboratory settings are very simple tasks, such as change detection tasks or visual search tasks. By using simple tasks in cognitive experiments, the target module of cognition can be isolated and noises resulted from individual differences in other modules of cognition can be minimized. But on the other hand, however, those simple tasks used in the laboratory setting are so simple and so straightforward that they do not reflect the mechanism that makes human vision stand out from computer vision.

One new direction of research has recently emerged from studying the behavior of human complex visual problem solving. Since MacGregor and Ormerod's [47] showed that human can

find near optimal solutions of Traveling Salesman Problems without much effort, human behavior on visually presented optimization problems have captured the attention of a number of cognitive scientists [9-11, 47-50]. Because optimization problems are mathematically well-defined, objective measurement of human performance is often readily available. Meanwhile, finding the optimal solutions for some of those optimization problems is computationally hard. Even looking for a sub-optimal solution may require a significant amount of computational resources in terms of memory and time. Since optimization problems are both well-defined and computationally hard, they provides a platform for investigating the underlying mechanisms of human visual system that support complex visual problem solving behavior.

2.4 PREVIOUS MODELS OF TSP

I begin by reviewing common models previously proposed for human performance on the TSP.

1. Nearest neighbor

The most basic model of TSP is the nearest neighbor model [51] in which the problem solver always selects the closest next point to the current point, i.e., simply following a hill-climbing heuristic. The model is elegant in that it only assumes a single heuristic that is already known to be part of the human information-processing repertoire [17]. However the model makes no use at all of global information and tends to produce solutions that are not as good as those found by humans [52].

2. Convex hull

The next simplest model of TSP is the convex hull model [53], which assumes that people compute a traversal around the perimeter points, including inner points opportunistically along the way using a minimal insertion rule. The global information used by this model is the Convex Hull contour, which may be rather complex, and thus require significant working memory. The minimal insertion rule is applied globally at each point in time during path computation, and points are added that cause the smallest increase in total path length. It is somewhat implausible that people would be able to compute these minimal insertions (a local processing task) at the global level due to the visual working memory constraint.

3. Sequential convex hull model

MacGregor et al. [50] adapted the convex hull model to a more plausible incremental local search version of the convex hull model. In support of this adaptation, they

found that humans perform better on problems with fewer interior points within the convex hull [47]. Second, their experiments provided support for their hypothesis that human participants are sensitive to global information [52]. I would call this model the sequential convex hull model. The outline of the model is as follows [50]:

3.1. Sketch the connections between adjacent boundary points of the convex hull.

3.2. Select a starting point and a direction randomly.

3.3a. If the starting point is on the boundary, the starting node is the current node. The arc connecting the current node to the adjacent boundary node in the direction of travel is referred to as the current arc. Proceed immediately to Step 3.4.

3.3b. If the starting point is not on the boundary, apply the insertion rule to find the closest arc on the boundary. Connect the starting point to the end node of the closest arc, which is in the direction of travel. This node becomes the current node.

3.4. Apply the insertion criterion to identify which unconnected interior point is closest to the current arc. Apply the insertion criterion to check whether the closest node is closer to any other arc. If not, proceed to Step 3.5. If it is, move to the end node of the current arc. This becomes the current node. Repeat Step 3.4.

3.5. Insert the closest node. The connection between the current node and the newly inserted node becomes the current arc. Retaining the current node, return to Step 3.4 and repeat Steps 3.4 and 3.5 until a complete tour is obtained.

4. Pyramid model

Pyramid model [9] of the traveling salesman problem was inspired by a hierarchical architecture of human visual and spatial perception. Their model first Gaussian-blurs the original set of points into a variety of degrees and stores those blurred images in different layers of hierarchy with the most blurred image on the top. The more blurred images serve as the global information for the less blurred images. Each layer directly guides the next layer below it each time the model develops a node into the path. So layers in the hierarchy change in a repeatedly cascaded process. The pyramid model computes TSP solutions in the following steps:

4.1. Gaussian-blur the original n-points TSP image into k-1 different degrees and store them in a k-layer pyramid with the original TSP image on the bottom and the most blurred image on the top.

4.2. Calculate L_i centers of the image in each layer i. Consider those centers in each layer as nodes in a reduce-sized TSP problem. The top layer has 3 nodes and the bottom layer has n nodes. Layer k has $\frac{n}{b^k}$ nodes. (The parameter b is the reduction ratio. Bottom layer is layer 1.)

4.3. Layer n (top layer) has 3 nodes and forms a unique tour.

4.4. Generate a tour of the TSP in each layer by inserting them into the tour on the previously higher layer with the following rules: (a) Sort the intensity level of the mode locations in each layer. (b) Insert these modes into the tour in descending order of their intensity, so as to produce the minimum increase in tour length. Repeat Step 4.4 until the algorithm generates a tour in the bottom layer.

This pyramid model provided another explanation on how human are able to find short paths through many points, but like other heuristics the VWM constraint were not formally taken into account in the model.

2.4.1 The missing role of visual working memory in complex visual problem solving

Despite all the research in visual working memory and on complex visual problem solving, few studies have investigated details in how visual working memory supports complex visual problem solving behaviors. One reason for this gap of research is because that the contents of visual working memory can't yet be directly monitored, although recently brain-imaging and machine learning technology have shown to be promising to do this in the future [54]. In the next few chapters of this dissertation, four experiments were reported, seeking the answers for the following three questions:

First, what kind of role does global information and external representations plays in complex visual problem solving?

Second, how is visual working memory allocated to global and local information to support complex visual problem solving?

Third, how does global and local information share the capacity limitation of visual working memory?

3.0 HUMAN COMPLEX VISUAL PROBLEM SOLVING AND GLOBAL INFORMATION PROCESSING

Except the nearest neighbor heuristic, all the other models of human TSP solving defined some sort of global information to support the local reasoning in the problem solving process. From the visual working memory constraints perspective, it is also intuitive that some sort of global information must be involved. If everything kept in the visual working memory are the positions of individual points of the TSP, then the only max amount of information taking into account at each step are 2-6 points. In that way it will be very easy for the problem solver to stuck at local optimal like the nearest neighbor heuristic, since the decision must be made by using less than 10% of all the points for a TSP with 60 points or more. So to study what kind of global information is used and how it affects human TSP solving, I did the first experiment.

3.1 EXPERIMENT ONE

3.1.1 Method

Participants

Twenty-eight graduate students participated in the experiment.

Materials and methods

The materials were 20 TSPs. Ten of them are real world problems borrowed from TSPLIB (http://www.iwr.uni-heidelberg.de/groups/comopt/software/TSPLIB95/) ranging in size from 16 points to 100 points. The remaining 10 were randomly pre-generated according to a uniform distribution ranging from 10 to 80 points. Note that all participants saw the exact same 20 TSP problems but in a random order, which allows us to examine how well the models predict the performance on particular TSP problems rather than just general trends for the effect of number of points. The problems were displayed in an 800 * 800 pixels window on a 17-in. computer screen with resolution 1440 * 900 pixels. Participants sit about 17-20 in. away from the computer screen. So all the problem lies in the human visual field with maximum angle of 10^{o}-13^{o}. Participants were asked to find the shortest possible path and indicate the path using mouse-clicks. The program recorded all the click data. Participants were randomly assigned into one of the following three groups. The groups were designed to examine the influence of the global and local information.

Control (10 participants)

Each participant was asked to solve the TSP problems while all point locations remained on the screen throughout.

Global preview (9 participants)

Each participant was asked to solve the same TSPs as in the control condition, with three distinct phases for each TSP.

1. The full TSP is shown, but paths cannot yet be clicked. Each participant was given a pen and a piece of paper to draw the global information they would need in the later phases. Participants were also asked to pick a start point to begin their TSP trip (Figure 3-1).

2. The TSP problem points were clustered into 5-12 clusters using a K-Means algorithm [55].The k-means centroids (geometric centers) were displayed as larger dots. Participants were asked to pick a path through just the centroids to determine the order in which the clusters show up in phase 3 (Figure 3-1).

3. All points were hidden. Then subsets of points were presented one cluster at a time, and participants had to pick a path through all the points within a cluster. When all the points in the current cluster were visited, the next cluster of points would become visible (Figure 3-1).

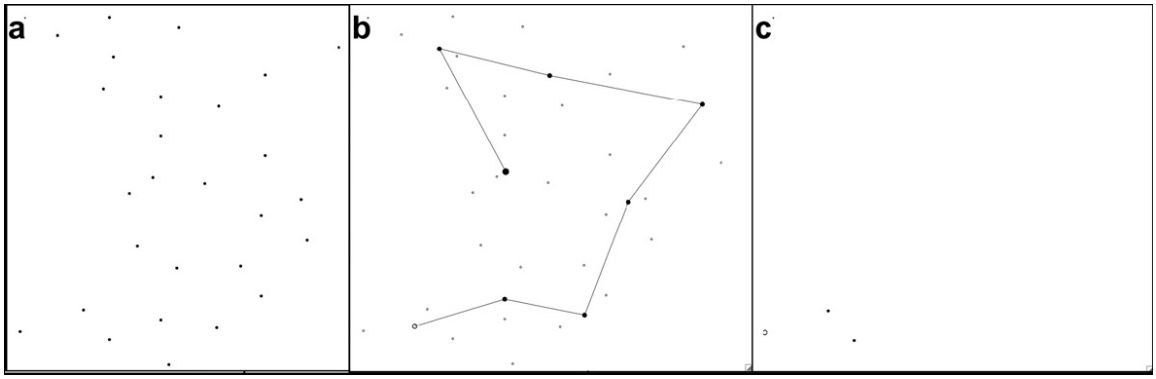


Figure 3-1. Illustrations of global-preview condition phase 1, 2, 3(left, mid, right).

No global preview (9 participants)

This condition was identical to the Global Preview condition in that only one cluster worth of points is shown at a time during the path selection process (i.e., phase 3), except that participants did not first see the full set of points (i.e., phases 1 and 2 were skipped). So there was no global information available during any part of the process.

The No Global Preview vs. Global Preview comparison tests the effect of access to global information on local search quality, and whether the global information can fit in working memory (as opposed to it being important that global information be externally available). If those two conditions do not differ in solution performance, then the control condition assesses whether even continuously available global information is helpful. If the Global Preview and No Global Preview conditions do differ, then the Control

condition assesses whether to what extent continuously available global information further shapes local search. Finally, the Control condition also provides baseline TSP

problem solving data against which the computational models can be compared. After each participant finished all 20 TSPs, there was a post-experiment measurement on how fast the participant clicked the mouse. This step involved re-presenting all 20 TSPs, but instead of finding the shortest path, participants were asked to click through all the points as fast as possible in an arbitrary order. From this data, I will estimate participants' thinking time by subtracting mouse-clicking time from solution time.

3.1.2 Results and discussion

Accuracy and reaction time were calculated as our measurements of performance. Accuracy was calculated as the ratio of participant path length over the optimal path length. Reaction time was

calculated as difference between the time to finish the TSP and the time to click through all the points. So accuracy is a number larger than 1. The closer the value is to 1, the better the performance is. The reaction time is an approximation of participant thinking time. One-way between groups ANOVAs on accuracy and reaction time revealed significant effect on both accuracy ($p < .0001$) and on reaction time ($p = .0001$) between groups. But the condition effect of reaction time ($F(2,8) = 9.0$) is much weaker than that of accuracy ($F(2,8) = 172.1$), while they have the same degree of freedom. The control group had the best accuracy 1.05 but highest RT 76s. The global preview group had middle levels on both (accuracy = 1.11, RT = 54 s). The No Global preview group had the worst accuracy 1.16 but fastest RT 42 s. Post-hoc Tukey comparisons found significant pair-wise difference between all groups on accuracy ($p < .0001$). That the control condition is significantly slower than the no global preview condition ($p < .0001$) and the global preview condition ($p = .0074$) suggests that processing global information does take time.

That the condition effects are very strong on accuracy ($F(2,8) = 172.1$) and much weaker on RT ($F(2,8) = 9.0$) suggest that a simple speed-accuracy tradeoff could not explain the overall condition effect. As it can be seen in Figure 3-2, the accuracy of each group slowly goes up (less accurate) when the size of the problem goes up.

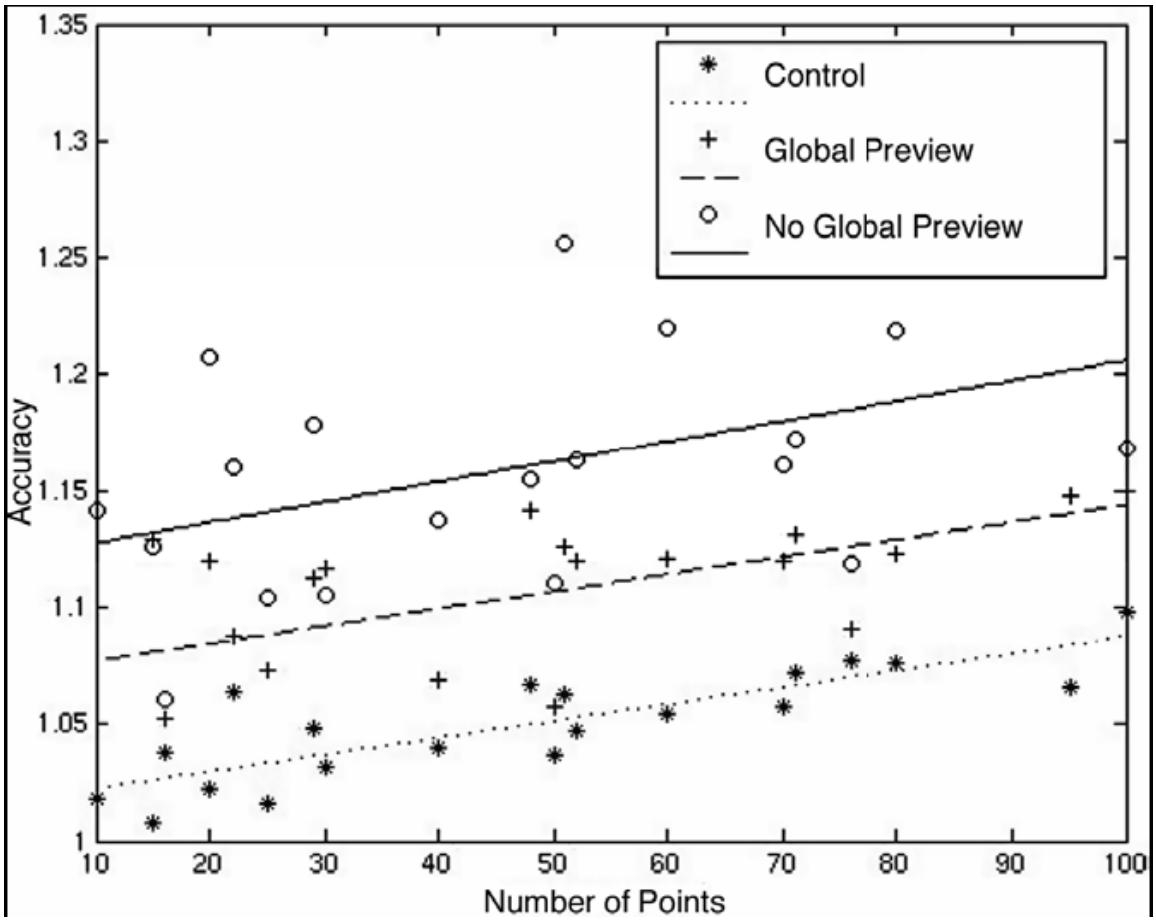


Figure 3-2. Mean accuracy of each individual problem within each condition

The accuracy of control condition fits well to a linear trend (R^2 = .689). The accuracies of the other two conditions basically follow linear trends (R^2 = .33 and .21). The control group has the highest accuracy performance. This result is consistent with our hypothesis that human participants utilized both global information and local information to solve the problem. The control group has all points visible during the entire problem solving procedure; the points on the screen appear to help them to retain the global information through some kind of active memory during the solution process. The global preview group has better accuracy than the group w/o global preview. This result confirmed the importance of global information in the human TSP

solving procedure. A typical example of the scratch notes of participants in the global preview group is in Figure 3-3.

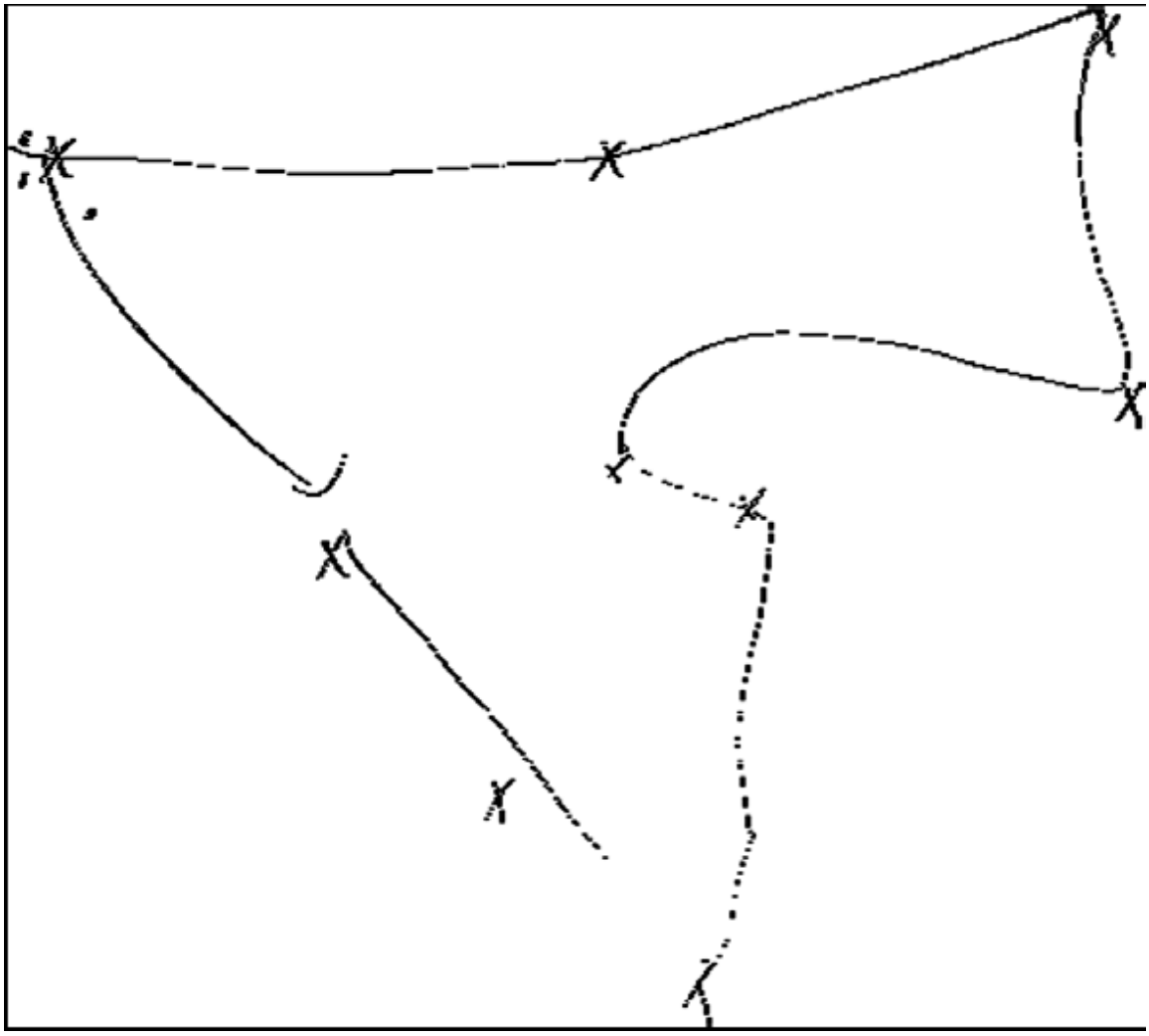


Figure 3-3. A typical scratch note from a participant in the group with global preview

Not all participants sketched a Spline-curve. Some participants just recorded the relative position of each cluster and some just left the scratch paper blank. But when connections between clusters were drawn, they tended to resemble splines. In sum, it appears that global information stored only mentally does help local search. Global information presented throughout problem solving helps even more. Thus, global information computed in global preview condition, either slightly

exceeds capacity limits, and/or is not stored with the same fidelity as global information that is supported with continual visual input. One could interpret the results as a support that human TSP problem solving relies heavily on compact global representations suggested by the notes of global preview condition. However, one could also argue that human TSP may use elaborate global information that is dependent upon constant peripheral visual input, since control condition has better accuracy than global preview condition. So the question is whether the recorded global information in global preview condition is close to the one used in control condition or it's only an abstract of it. Precise modeling of the exact human data in the control condition may help to resolve this.

3.2 OUR FIRST MODEL OF HUMAN TSP SOLVING

Based on this theoretical analysis and observations of human behavior in the global preview condition, I propose our first model for TSP problem solving: K-Means TSP model. Our K-Means TSP model is based on the following three steps:

1. Clusters are identified.

In this step, points are grouped according to visual density. Points constructing a higher visual density are more likely to be grouped together. Our model approximates this clustering identification process using a K-Means clustering algorithm, because it is available in standard software packages. The K-Means Clustering Algorithm clusters N data points into M disjoint subsets S_j containing N_j data points so as to minimize the sum of squares criterion:

$$J = \sum_{j=1}^{K} \sum_{n \in S_j} | x_n - \mu_j |^2$$

where x_n is a vector representing the nth point and μ_j is the geometric centroid of the points in S_j. All the centroids are added into the collection of reference points in VWM, which was passed from the previous iteration.

2. A sketch of the path is conceived.

Here by sketch of the path, I mean the path visiting all the groups and returning to the starting group. Using this strategy, human cognition reduces the original problem to a main problem of much smaller size with simple sub-problems. Here I use a Spline-curve of all the centroids to model this sketched path.

3. Connect all the points along the sketched path.

I model this step using a projection rule. I project all the points to nearest point on the Spline-curve. Then I construct the final solution by connecting all the points in the same order as their projection on the Spline-curve. Steps 1 and 2 of our model are the global information processing part, and Step 3 is the local information processing part. The global information perceived in Steps 1 and 2 will guide the local information processing in Step 3. The Spline-curve is the global information developed after Steps 1 and 2. The clusters and centroids are no longer needed after the Spline-curve is sketched. So in the local search phase, the cluster and centroids information can be discarded, since the Spline-curve itself is enough to guide the local information processing in Step 3. The Spline-curve plotted fits both of our criteria for global information. First, it is sufficient to guide the local search in the third step of the model, where the model only need to project the points onto their nearest curve. Second, because clustering result and centroids information can be discarded after Step 2, the Spline-curve itself is compact

in size and has a visual representation that may fit well to human visual/spatial working memory capacity. Our hypothesis is that there are some visual operators for human cognition that enable it to do the first two steps within a near constant time and the third step in a linear time. Figure 3-4 illustrates the three steps of our model when solving a 70-points TSP.

Model simulation

I used a fixed the k-means centroids in the upper right plot set of 20 problems across participants in our experiment. The negative consequence of this experimental design choice is that I do not have a pure estimate of the effect of problem size because of the small idiosyncrasies of our chosen problems. However, the positive consequence is that I have enough data for each exact problem to evaluate how well each model can explain performance on those particular problems, in trends across problems and exact fit to problem performance. The number of clusters is the only parameter setting in this model. In our simulation I set it to $2\sqrt{n}$, where n is the total number of points in the TSP. This setting was based on the intuition that I do not want points to be too far away from its cluster centroid to avoid too much error. If recursion has to happen for large TSPs, I want the depth of recursion to be no more than one. In order to draw the Spline-curve around all the centroids in step two of the model, I need a TSP path around all centroids. In our simulation I used a recursive call to our model until the size of the problem is below 6, when I can easily use an exhaustive search function to find the shortest path around the remaining points. For our current set of problems, the depth of recursive calls is at most two. Since the K-Means clustering algorithm may converge to a local minimum and may yield different clustering results on different runs, I ran through our model on all the 20 problems 40 times.

The mean accuracy of the 40 runs is pretty close to the accuracy performance of participants in the global preview group. Since our model employed a naive local search strategy and the group

with global preview had incomplete local information, the closeness of accuracy between them is what I expected. However, the minimum accuracy generated by the 40 runs of our model is very close to the accuracy performance of the control group. The reason might be sometimes the naive projection rule in our local search fits to the generated global information very well, so it produced a similar result as the more sophisticated local search strategy used by human cognition.

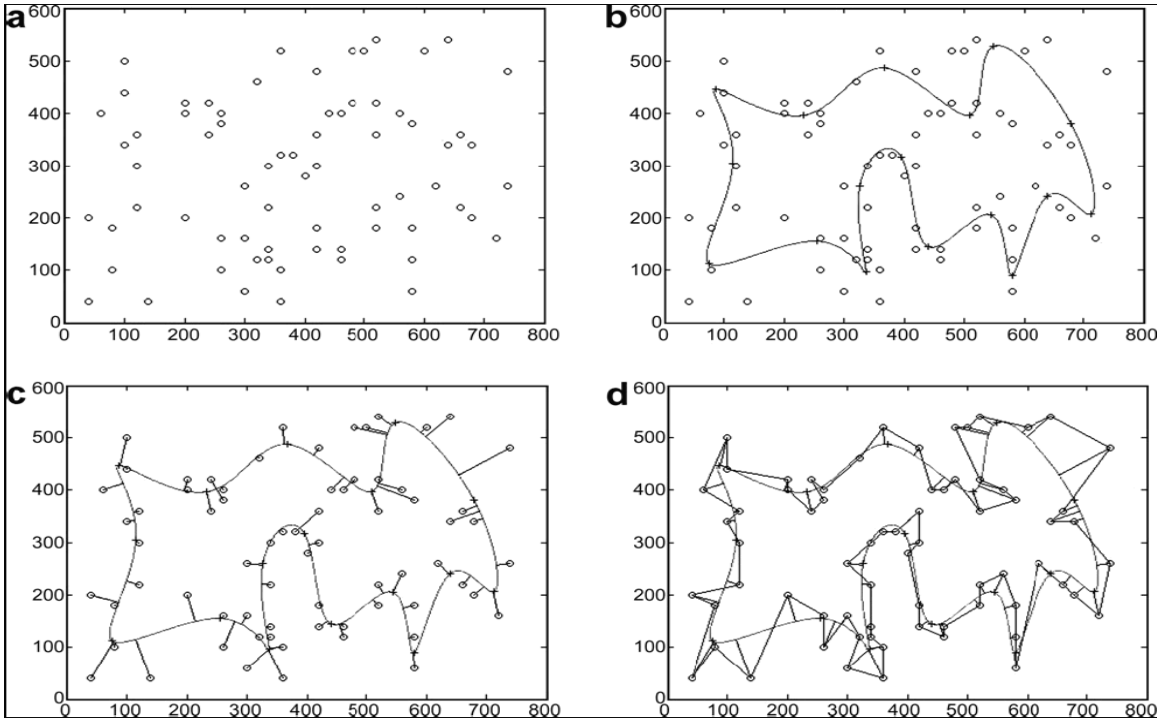


Figure 3-4. Three steps of the Kmeans-TSP model on solving a 70 points TSP. "+" indicate the locations of the Kmeans centroids in the upper right plot.

3.3 EVALUATION OF OUR FIRST MODEL

3.3.1 Number of intersections

For human and models, I computed the number of times the selected final path crossed itself (called intersection). For data from humans and models with a random factor (Human, NN, CHSQ, Kmeans), I computed means (see Table 1).

Table 1. The sum (over all 20 problems) of means (over different participants or model runs) on number of intersections generated. Correlation and ASE on number of intersections between each model and human participants; "*" indicates best matched model on each dimension.

	Human	NN	Convex Hull	Pyramid	Kmeans	CHSQ
Number of intersections	2.3	110	6	14	2.8	2.3
Correlation		0.30	−0.13	0.23	0.31	0.39 (*)
Average signed error		4.98	0.15	0.63	0.03 (*)	−0.03 (*)

Figure 3-5 plots the number of intersections generated by human and each models on the 20 TSP problems used in our experiment. Because NN has no global information, it generates many more intersections than the rest of the models and human data. Pyramid and convex hull have deterministic algorithms, so they generated certain high peeks on particular problems and zero values on others. Both Kmeans and CHSQ are close to human data in value of number of

intersections. None of the correlations with human performance were statistically significant, although the CHSQ correlation was marginally significant ($p < .1$).

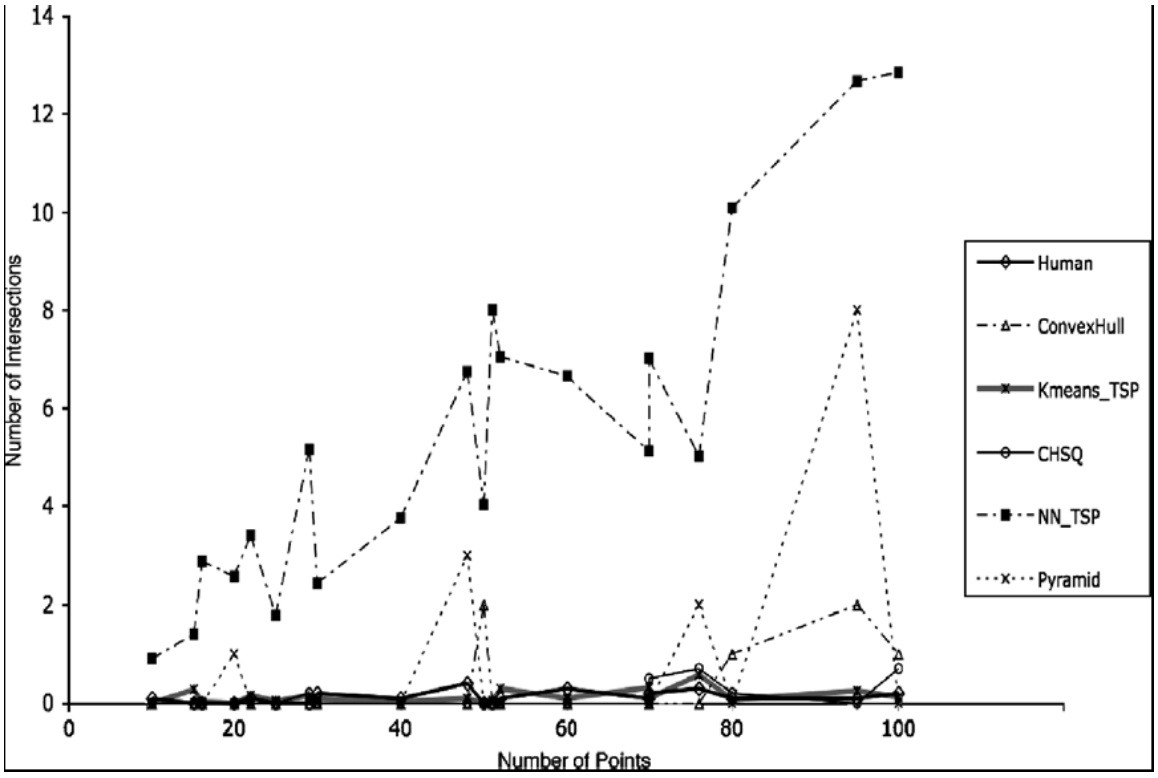


Figure 3-5. Mean count of intersections generated by each model for each problem

3.3.2 Mean accuracy

As shown in Table 2, NN is much worse than the human performance in term of mean accuracy. Convex Hull, Pyramid, Kmeans and CHSQ are close to human accuracy levels as has been found in the part. Kmeans model did depart from the human data and the other three models as the

number of points got larger. One reason for this is Kmeans model employed a naïve local search strategy that project points onto their nearest Spline-curve. As the number of points going bigger, the ratio of centroids to points is smaller. So the Spline-curve is more inaccurate in characterizing the detail local information. In this situation a more sophisticated local search strategy should be employed.

Table 2. Correlation and ASE on accuracies between models and human participants. ("*" indicates best matched model on each dimension.)

	NN	Convex hull	Pyramid	Kmeans	CHSQ
Correlation	0.70	0.62	0.41	0.84 (*)	0.67
Average signed error	0.17	−0.01	0.02	0.04	0.00 (*)

All but the Pyramid model led statistically significant correlation with (human data). The Kmeans model correlated with the trend of human performance best among the models I compared. Our hypothesis is that the global information Kmeans model utilizes is the best approximation to the global information human use, so it generate a similar trend with human performance. Figure 3-6 plots the means of accuracy of human performance and each model.

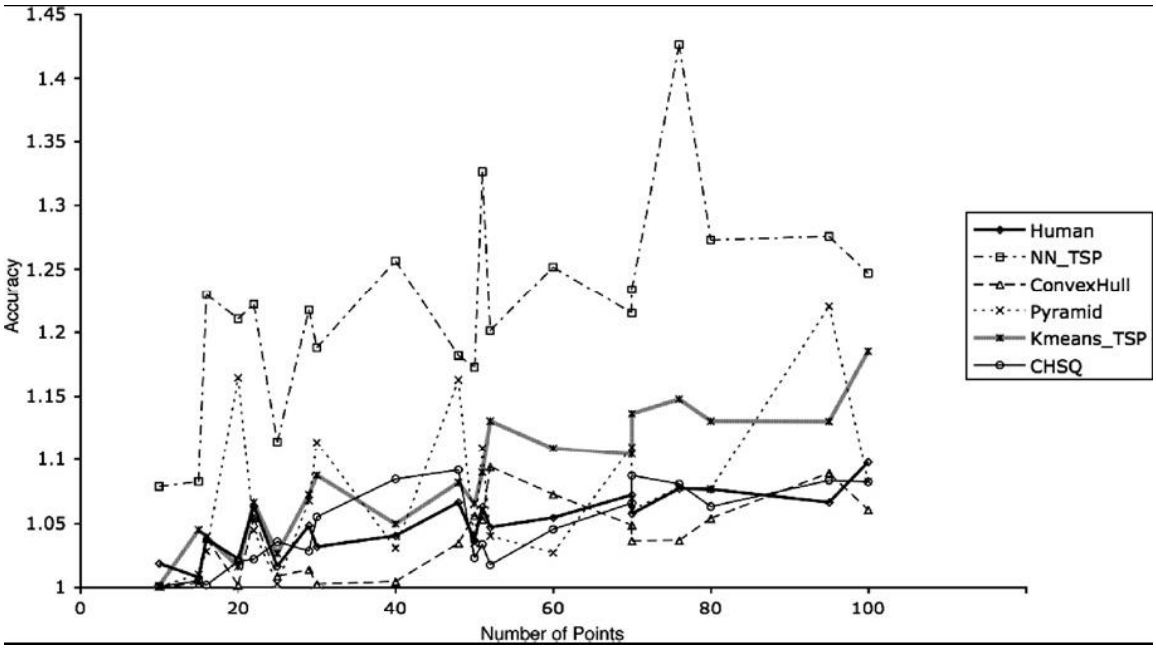


Figure 3-6. Accuracy performance of models and humans on each problem.

3.3.3 Standard deviation

In addition to accounting for overall and problem-specific differences in mean accuracy, a model could also try account for overall and problem-specific differences in the variability across participants in accuracy (as measured by standard deviations). These differences in variability might reflect the degree of garden path effects from different start points (i.e., small choices made early have large down-the road consequences). A Levene Test shows that there are statistically significant differences in the standard deviations of human accuracies on different problems.

At the level of overall standard deviations, the pyramid and Convex Hull models fail outright because they are deterministic, and thus predict standard deviations of zero. The nearest neighbor model predicts standard deviations that are too large. The Kmeans and CHSQ models are close

to observed human levels overall (Figure 3-7). In terms of predicting problems specific differences in variability, none were statistically significant. It may be because that a few participants who were using different strategies than others (see Table 3).

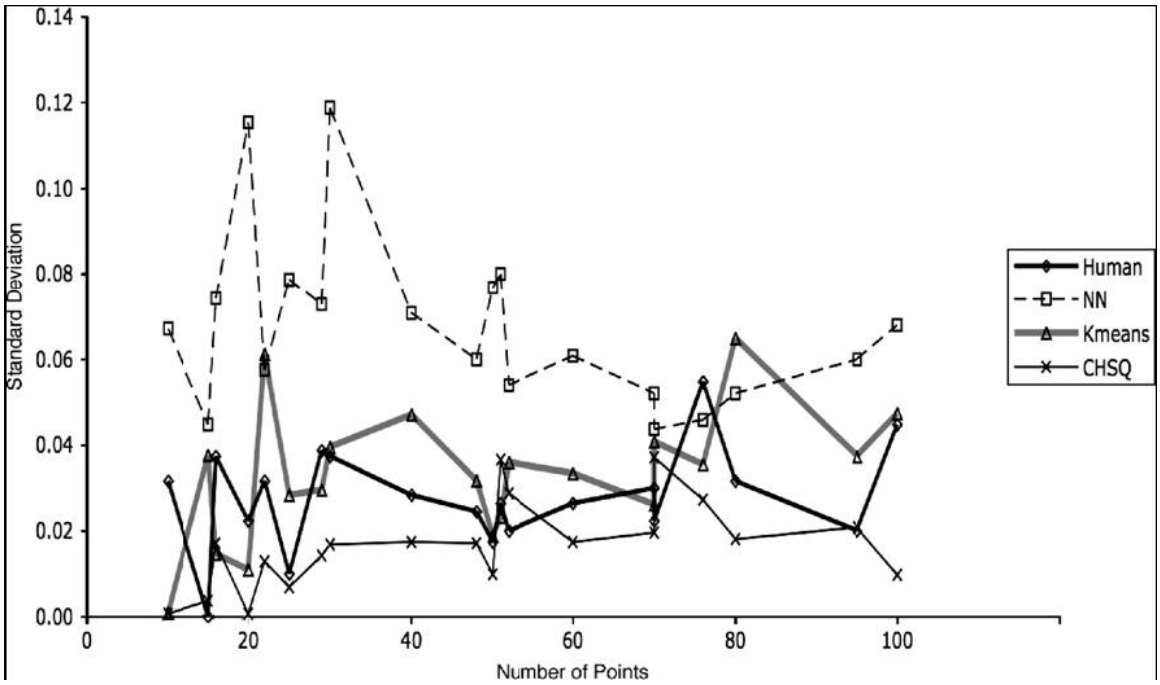


Figure 3-7. Standard deviation on accuracy of human participants and the three models with any variability.

Table 3 Mean standard deviation, correlation and ASE of the standard deviations between human participants and models. ("*" indicates best matched models.)

	Human	NN	Kmeans	CHSQ
Mean	0.03	0.07	0.03	0.02
Correlation		0.06	0.11	0.22

	Human	NN	Kmeans	CHSQ
Average signed error		0.04	0.01 (*)	−0.01 (*)

3.3.4 Exact path correlations

A good model of human TSP problem solving should not only predict the accuracy of the total path length that a human would generate on a TSP problem but also should be able to predict the likelihood of human participant taking a particular path. I used the following method to calculate the exact path correlation between human-generated and model-generated solutions. For each TSP problem with n cities, build a matrix of n × n, where each cell M(i, j) equals the numbers of observed paths between city i and city j. Then compare the similarity between the models and the participants at the individual path level by linearizing the matrix and compute the correlation between the two resulting vectors. Figure 3-8 shows the mean correlation between the participant solutions and model solutions on each problem.

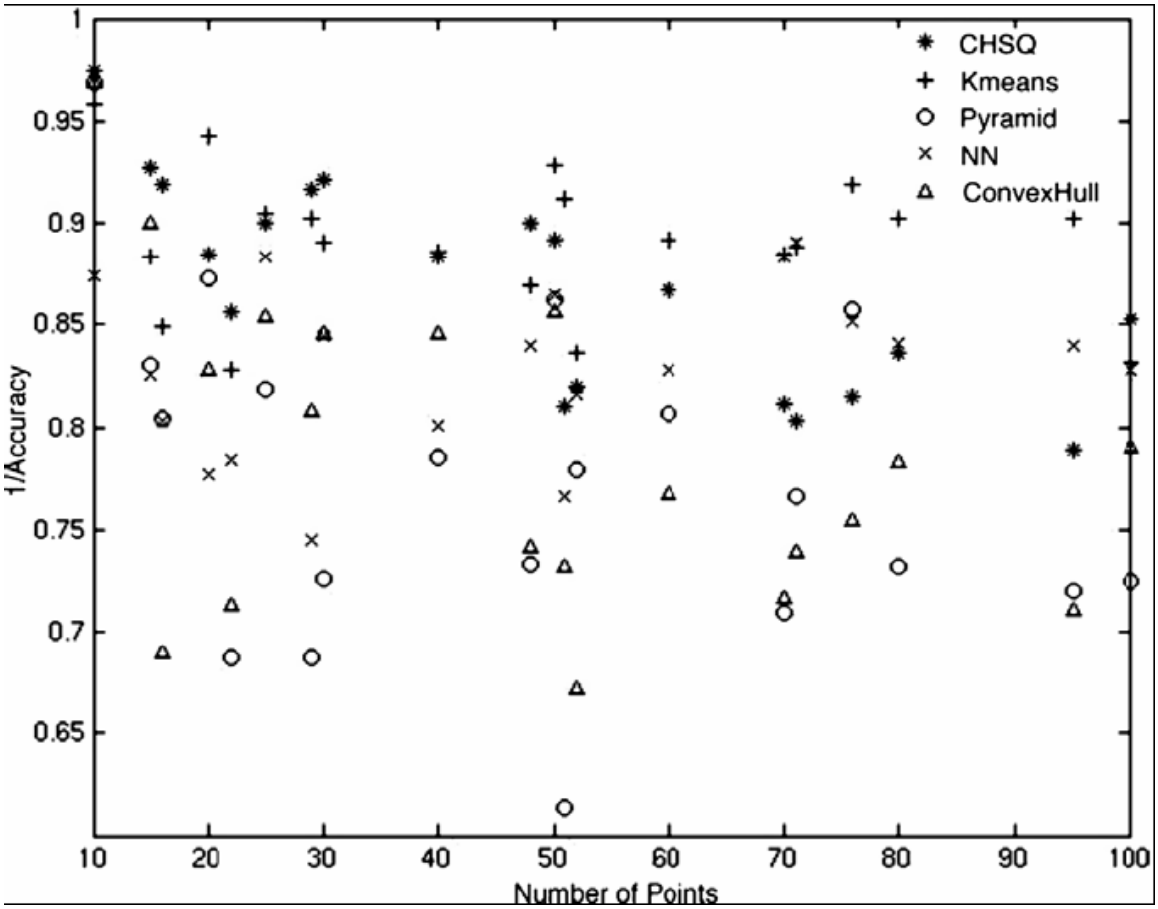


Figure 3-8. Exact path correlation between human participants and each model.

As Table 4 shows, the Kmeans model correlates best with the participants' choices on exact paths. The Kmeans model outperforms other models especially on larger size problems ($n > 50$), as Figure 8 shows. One possible reason for this phenomenon is that as the size of the problem grows larger, human participants display a larger diversity of possible path choices. Our model captured this characteristic of human TSP solving by generating different paths on each run.

Table 4. Mean correlations between participant and model solutions on each of the 20 problems. ("*" indicates best matched models.)

	NN	CH	Pyramid	Kmeans	CHSQ
Mean correlation	0.83	0.78	0.77	0.89(*)	0.87

I visualized the characteristic by plotting the frequency of an edge selection by participants or each model as its thickness. Figure 3-9 shows the solutions of a 50-points TSP generated by participants, Kmeans, CHSQ and Pyramid models. The arrows point to areas where participant-generated paths and Kmeans-generated paths displayed a great similarity in both pattern (path choices) and thickness (path frequency). The same kind of similarity could not be found in other models.

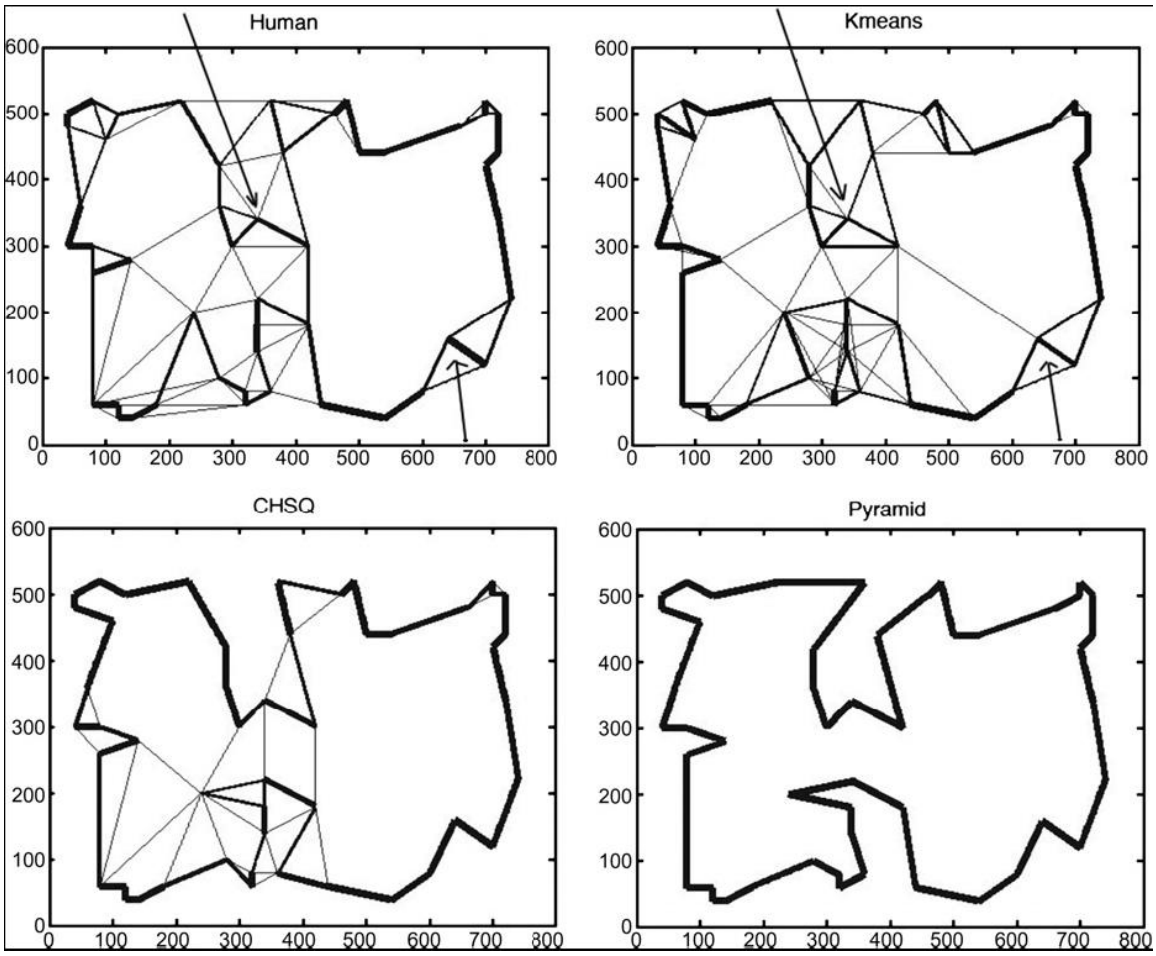


Figure 3-9. Chosen paths for a 50-point TSP generated by participants and models.

This visualization technique also helps us to identify reasons for why the Kmeans model departed from human data. As I can see in Figure 3-10, the main outside contour of the Kmeans model displayed a large number of zigzags (as the arrows identifies) while the participants, the Pyramid and CHSQ models did not. The zigzags in the Kmeans model are the result of its naïve local search rule of projection. Since the Kmeans model connects points that have the nearest projections on the Spline-curve, those points themselves could be far away if one point is inside

the Spline-curve and the next one is outside it. Through repeated crossing of the Spline-curve, the zigzag pattern is generated.

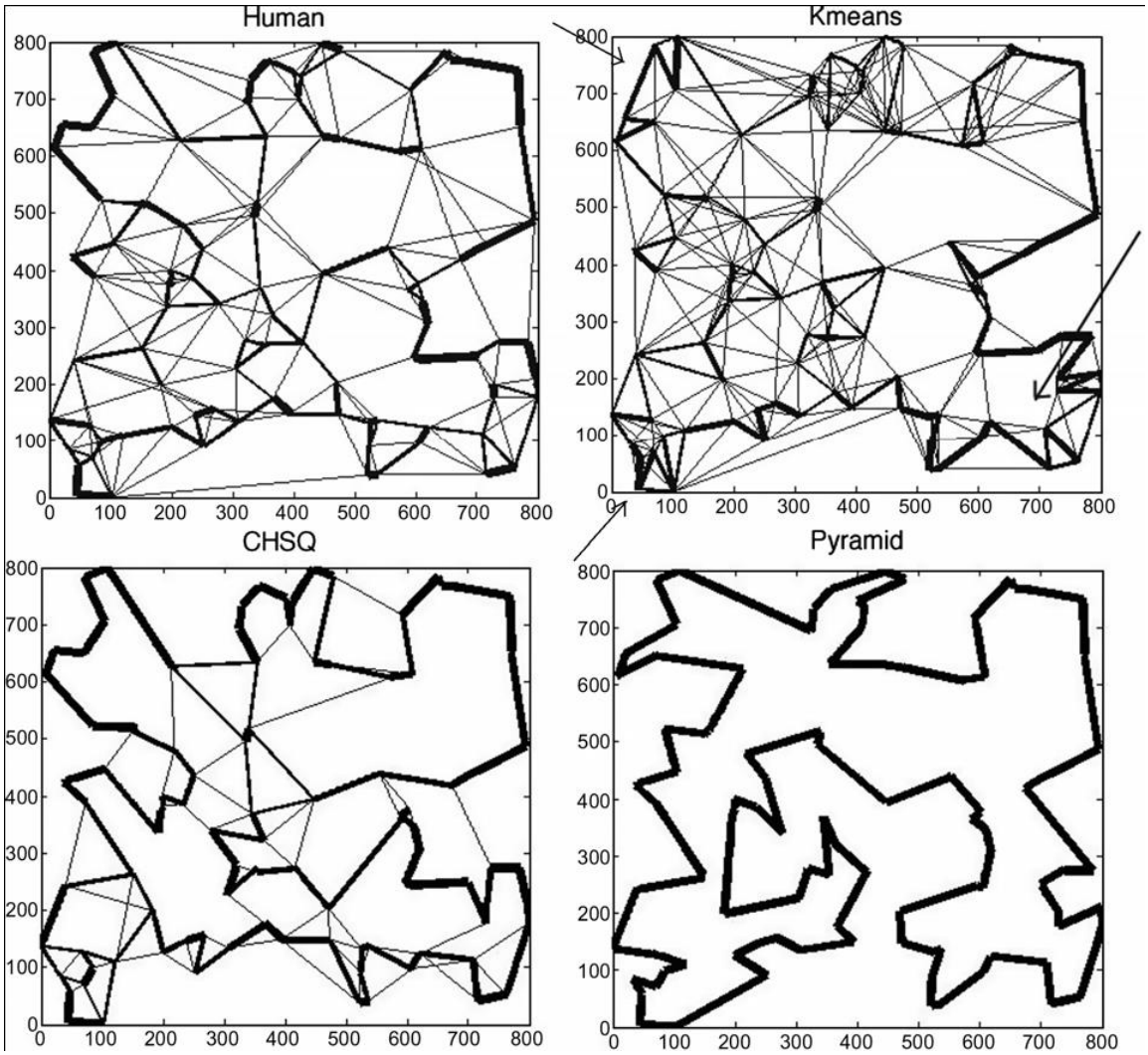


Figure 3-10. Chosen paths for a 100-point TSP generated by participants and models.

4.0 VISUAL WORKING MEMORY ALLOCATION IN COMPLEX VISUAL PROBLEM SOLVING

To seek an explanation of the paradox that how could such a limited visual working memory resource support powerful human visual problem solving abilities, I explored the important role of global information and introduced our first model of human TSP solving integrating global information processing in the last chapter. However, when deciding which point to visit first and which to visit later, one obviously has to take into account the local information as well, which is the exact location of the points. If one has to take into account both global and local information to make a decent decision when solving complex visual problems, then both global and local information has to be stored in the visual working memory simultaneously. If visual working memory stores both global and local information, the immediate question is how this very limited resource is allocated to store both type of information. Since there are only several slots of visual working memory, its allocation mechanism must be highly efficient. However, to study the visual working memory allocation mechanism straightforwardly requires keeping track of the contents in the visual working memory during each moment. Unfortunately, such a technology is still yet to be invented. But on the other hand, what I can measure is the locations of attentional fixations that encode visual information into VWM. As a recent study has shown, VWM

resources are allocated by attentional fixations and the contents in VWM are updated through eye-movements [32]. So to study the VWM allocation mechanism during complex visual problem solving, we can take a "reverse engineering" approach. Suppose we want to know the formula of Coke, but we don't have a method to directly measure its ingredients. But we have some hypotheses of what might be in Coke, we can try to make some drink with our hypothesized formula. If our drink tastes the same as Coke, then we could say that our formula is likely to be the same as Coke, because ingredients decide the taste. In the same way, since visual working memory allocation mechanism decides the eye-movement pattern, we can build a cognitive model to simulate the hypothesized visual working memory allocation mechanism and test it with human eye-movement data during complex visual problem solving using two experiments: Experiment 2 and Experiment 3.

4.1 OUR SECOND MODEL OF HUMAN TSP SOLVING: A MODEL ACCOUNT FOR THE DYNAMICS OF VISUAL WORKING MEMORY ALLOCATION

The following principles of VWM allocation were implemented in our model:

Principle One: The human visual system can represent visual information at different levels of detail. Each part of this information, regardless of its level of abstraction, is kept in one slot in VWM [56].

Principle Two: Information at the most global level is encoded first. This information then guides which local part of visual information should be focused upon and be further expanded into local fine information [57-59], which is also consistent with what the neural

measures of visual cortex suggested: global information is encoded first and guides the perception of refined information [60]. While attending to a local part, one chunk of global information is usually expanded into 2 or more chunks of local information.

Principle Three: Visual information most relevant to the current part of the task is represented at the most local level. Less immediately relevant visual information can be stored globally in VWM without being expanded. During visual complex problem solving, in order to keep the big picture in mind and focus on key details at the same time, both kinds of information are represented in VWM simultaneously.

Principle Four: Due to a capacity limit in VWM, some global information perceived previously will be lost. But a certain amount of global information is maintained in VWM all the time. Whenever the amount of global information in VWM runs low, attentional fixations will be needed to re-attend to global information and reload them into VWM.

I argue that a VWM allocation mechanism based on those principles, even with only a few slots of VWM to deploy, can represent complex visual information in a temporal manner enabling it to focus on the details without losing the big picture during complex visual problem solving. Meanwhile these principles of VWM allocation also predict quantitatively different eye-movement patterns across individuals with different VWM capacities, as I found in the current experiments.

Those VWM allocation principles were built into a model to predict human performance and eye-movement during Euclidean Traveling Salesman Problem solving, a classic paradigm for studying complex visual problem solving. The Euclidean Traveling Salesman Problem (TSP) requires finding the shortest path to visit a set of points on a Euclidean plane, returning to the start. Finding the exact solution of TSP is a NP-hard problem. So even for today's fastest

computer, there is still no efficient algorithm to exactly solve large instances. Different from existing models of human TSP solving [9, 49, 50], our model takes a VWM capacity limit into account with VWM size as a parameter and a VWM allocation mechanism based on the above principles. A sketch of the model is as follows while the details can be found in the supporting online materials section:

In the model, global information is first perceived by clustering all points into N clusters, where N is the VWM size. Each cluster of points regardless of its size is represented as a chunk of information and occupies one slot in VWM where only the centroid location of those points is kept (Principle One). In the case of TSP solving, because the cluster containing the current point is most relevant to the next movement decision, our model refines it into sub-clusters (Principle Two, Three). The chunk of global information in the VWM slot containing the current point is now broken down into M chunks of refined information, where M is a number greater than 2 and smaller than N. Because VWM is limited in size and holds only N chunks of information, the chunks of global information least relevant to the current decision are lost and replaced by the just perceived information. This keeps the total number of chunks in VWM under the capacity limit. Each time VWM updates its contents, the contents are manipulated and sorted based on the centroid locations of the clusters that those chunks are representing, so that these centroids form a path of shortest length when both origin and destination centroid are specified. The first chunk of information is refined recursively until each of the first two chunks contains only one point, where the first chunk contains the current point and the second chunk contains the next point to be visited. After connecting to the next point, the model checks if there is enough global information in VWM to guide the next decision. When there is, the model starts to refine the global information in the first slot of VWM. Otherwise the model visually re-attends to the

global information at the most global level, by moving the eyes to more remote aspects of problem (Principle Four). When processing global information, I assume eye fixations will be found around the center of gravity (centroid) of cluster of points due to the global effect of saccadic eye-movements [61-64]. These saccades are required to adequately encode the locations of objects outside foveal attention.

4.2 STATISTICAL MEASUREMENT OF EYE-MOVEMENT

To compare the eye-movement patterns the model predicts to those produced by humans, I define the distance of each fixation as the minimal of the following two distances: fixation point to the last visited point and fixation point to the next to be visited point (Figure 4-1).

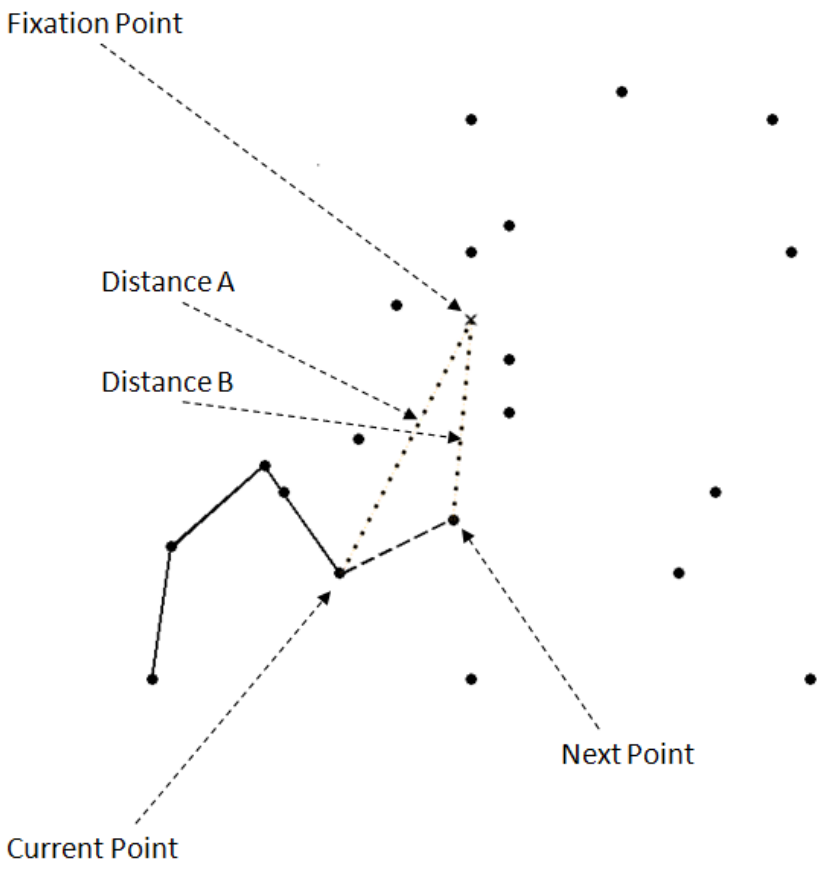


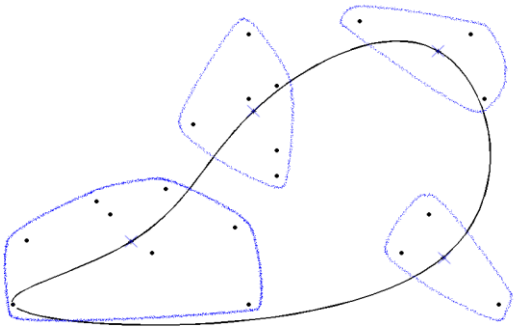
Figure 4-1. Fixation distance is the minimum of Distance A and Distance B.

Our intuition of this definition is as following: There are two types of fixations during this task. The first type of fixation is used to encode items into VWM for complex reasoning process before deciding where to go next. The second type of fixation is used to program motor actions to make a mouse click on the next to-be-visited point after one decides which point it is. Fixations of the second type are usually found near the point that is to be clicked. Before making those fixations, one has already shifted the current goal, since the next point has already been

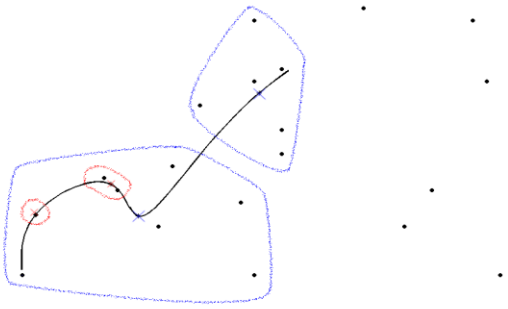
decided. Our definition of fixation distance measures how far away one's attention deviates from current goal regardless how far away the next to be visited point is from the last visited point.

4.3 MODEL PREDICTIONS ON EYE-MOVEMENT PATTERN

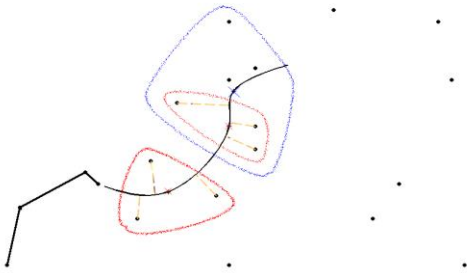
Generally our model predicts that fixation frequency decreases regularly as a function of fixation distance, according to the following logic. Each chunk of local information is an object consisting of only one or two points; while each chunk of global information may contain a cluster of points regardless of its size. Starting from the most global level, the cluster containing the current point will be expanded into finer information. If it were entirely expanded into local information, all global information would be lost. So the top level global information is only expanded into smaller pieces of the next level global information. The cluster of points that contains the current point is expanded recursively. As Figure 4-2 illustrates, the closer to the current point, the finer the information perceived and thus more fixations are required.



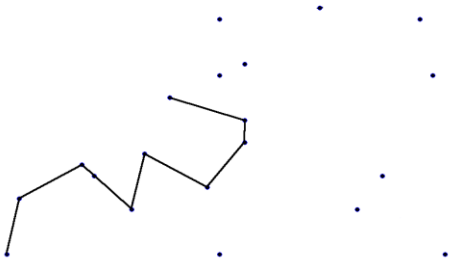
A



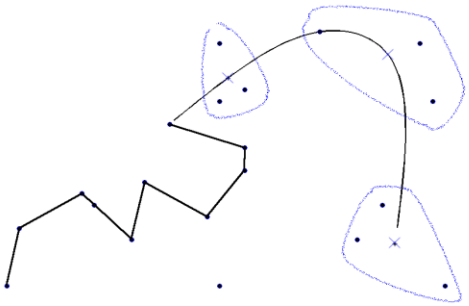
B



C



D



E

Figure 4-3. Steps in our second model of Traveling Salesman Problem solving

This figure illustrates a sample VWM allocation process during a TSP solving process.

Color contours are the objects held in VWM at each moment. Color crosses are the centroids of those clusters. Those centroids represent the locations of the objects indexed in the VWM.

A. Blue contours are the top-level global clusters. The blue crosses are centroids of those clusters. Each of those global clusters was treated as a single object and their location information was encoded in VWM. Our model makes attentional fixations on those blue crosses to encode the position of the each object. A rough sketch of the TSP path (black contour) is calculated based on global information.
B. Points between the current point and centroid of the next object were broken down into finer information (red contours). Top-level global information (blue contours) further away from the current point on the rough sketch were discarded from VWM due to VWM capacity limitation.
C. As the next chunk of objects on the sketch contour contains less than two points, the edge between the current point and next point on the sketch contour was connected.
D. At this point of the TSP solving process, local information was consumed as the edges were connected. There is not enough global information to guide the direction of next move.
E. Global information (blue contours) were re-acquired and encoded into VWM.

In fact, the model makes very precise predictions about the pattern on the relative frequency of eye fixations to increasing distances from the currently related points. The model also predicts that this pattern on the relative frequency of eye fixations will change quantitatively across individuals with different VWM size due to the following reason. As Figure 4-4 A-C shows, as global information breaks down into smaller pieces of finer information, global information less relevant to the current goal are discarded from VWM in order to store the finer information. When those pieces of finer information are consumed as the current point moves, one has to re-attend to farther aspects of the problem to re-acquire the top level global information into VWM in order to enable global planning (Figure 4-5 D). The first piece of global information is again broken down into finer pieces following the VWM allocation principles. So as the size of VWM increases, one will be going less frequently through this procedure of re-acquiring global information. Thus fixations with longer distances will be less found often.

I first tested the general prediction in an experiment with 11 participants.

4.4 EXPERIMENT TWO

4.4.1 Participants

11 undergraduate students volunteered to participate in the study. Gender of participants was not recorded.

4.4.2 Materials

For all participants, I used the same 20 TSPs as the ones used in our previous study [48]. 10 of them were randomly pre-generated according to a uniform distribution with size 10, 15, 20, 25, 30, 40, 50, 60, 70, 80. The other 10 of them were downloaded from an online TSP library "TSPLIB", which were used to test TSP algorithms and most of them are from real world problems. Those problems have 16, 22, 29, 48, 51, 52, 70, 76, 95, 100 points each.

4.4.3 Methods

Participants were instructed to use a mouse to solve 20 Traveling Salesman Problems (TSPs). Points of 0.1° visual angle size were displayed on a $30^{\circ} \times 30^{\circ}$ region of a 17-inch screen with a white background. A larger point (0.2°) was the starting point. From there, participants used a mouse to left click on the next point to visit, and a line is connected from last point to the one just clicked. The goal of the task is to find a path as short as possible to visit all the points and return to the starting point. (Figure 4-6) No training was given regarding what kind of strategies the participants should use to solve those problems.

During the experiment, there were 20 problems varying in size from 10 to 100 points. After each problem, performance feedback was displayed as a ratio of their solution length over the optimal solution length. A Tobii 1750 remote eye-tracker was used to track eye fixations during problem solving procedure. All the eye-movement data was recorded by the eye-tracker. All the mouse clicks data were stored by the experiment program in Matlab.

Participants were paid for up to $20 for this part of the experiment based on their performance---for every problem producing above or equal to average performance, the participant received $1.

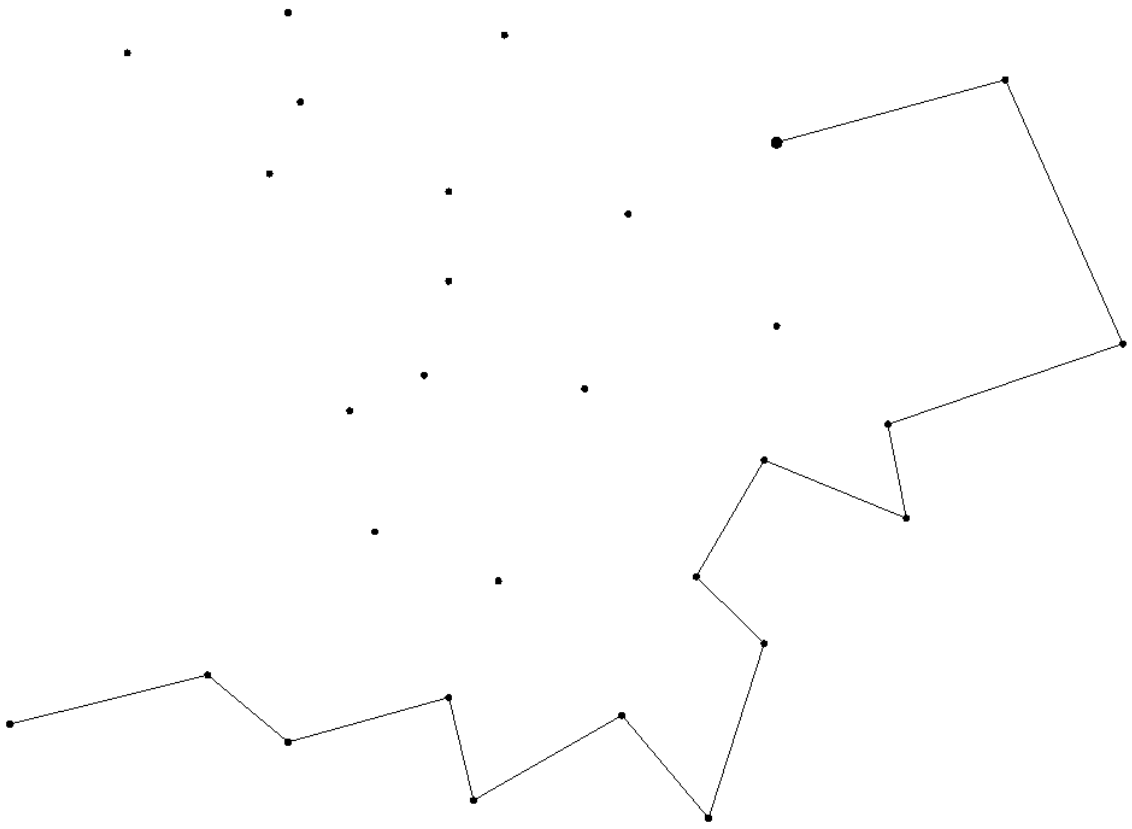


Figure 4-6 Example of a TSP solving trial. Participant left click on the next point to connect to it from the current one.

4.4.4 Results

They were asked to solve 20 TSPs while their eye-movements were recorded. Their fixation distances were extracted and were plotted as a histogram (Bins = 30). As predicted by the model, this histogram decreases along the fixation distance and is fit very well by an exponential curve (R^2 = 0.97). This finding provides preliminary support for our hypothesis that fine information was examined near the current goal and more global information was examined as distance increases.

I next investigated the relationship between eye-movement pattern and individual differences in VWM capacity. If our hypothesis were true, then different VWM capacities would generate quantitatively different fixation patterns. When VWM capacity is larger, more global information can be kept in VWM while local information occupies the other slots in VWM, so global information is less frequently re-attended. On the other hand, when VWM capacity is smaller, global information needs to be attended frequently in order to have enough information in VWM to make decisions. I ran our model with five VWM parameter settings (VWM = 3, 4, 5, 6, 7) each 40 times for all 20 TSPs. Each time an object (a cluster of points) is encoded in VWM, the model makes a fixation on the center of gravity on the object (centroid of the cluster). Then I plotted the histograms of fixation distances generated by the model.

Our hypothesized model predicts that the decay rate of the fixation distances histogram would increase as the capacity of VWM increase, because global information would be less frequently re-attended when there is enough room in VWM to keep it for a longer time. The type of curve that best fits the histogram goes from quadratic to exponential as the VWM size

increase from 3 to 7. Our model simulation predicts a trend in the decay rate of fixation distances as a function of VWM size. (Table 5)

Table 5. R^2 Fitness of different curves to model simulated fixations histogram (Bins = 30) and decay rate of the curve measure by the exponential factor b of the best fit exponential curve ae^{-bx}.

Curve Type / VWM	Quadratic	EXP.	Power	EXP. Decay Rate b
3	**0.982**	0.954	0.685	.0039
4	0.939	**0.969**	0.831	.0054
5	0.913	**0.970**	0.862	.0058
6	0.903	**0.972**	0.880	.0059
7	0.820	**0.984**	0.906	.0078

To test this prediction, I did the third experiment in which I first estimated 31 participants' VWM capacity and then asked them to solve the 20 TSPs. Participants' eye-movements were recorded during TSP solving.

4.5 EXPERIMENT THREE

4.5.1 Participants

31 volunteers non-color-blind participated in this experiment with age ranging from 18 to 40. Gender of participants was not recorded.

4.5.2 Materials and methods

Part One: Estimating individual visual working memory (VWM) capacity.
I used sample arrays consisting of 1 to 12 colored squares ($3^{\circ} \times 3^{\circ}$), each of which are selected at random from a set of 7 highly discriminable colors (red, blue, violet, green, yellow, black and white). All stimulus arrays were presented within a $30^{\circ} \times 30^{\circ}$ region of a 17-inch screen with a grey background. The positions of items were randomized in a given array with restriction that items were separated by at least 4.5° (center to center). One item in the test array was different from the corresponding item in the sample array by its color on 50% of trials; the sample and test arrays were otherwise identical. In each trial, the sample array stayed on the screen for 500ms. Then, after a blank screen for about 900ms, the test array was up on the screen for 2 seconds. Participants were instructed to press the "s" key if the test array was identical with the sample array or the "d" key if different within those 2 seconds. (Figure 4-7)

During this part of the experiment, there were 4 sections and each section has 60 trials. Participants were given opportunities to rest between sections. The experiment program in Matlab recorded all responses. Participants were paid $3-$7 based on their performance for this

test (if the visual working memory score was below 2.5, then $3; for every 0.5 addition score, another $1 was given).

Part one of the experiment was immediately followed by part two.

Part Two: Traveling Salesman Problem Solving

Part two of the experiment was a repetition of experiment one.

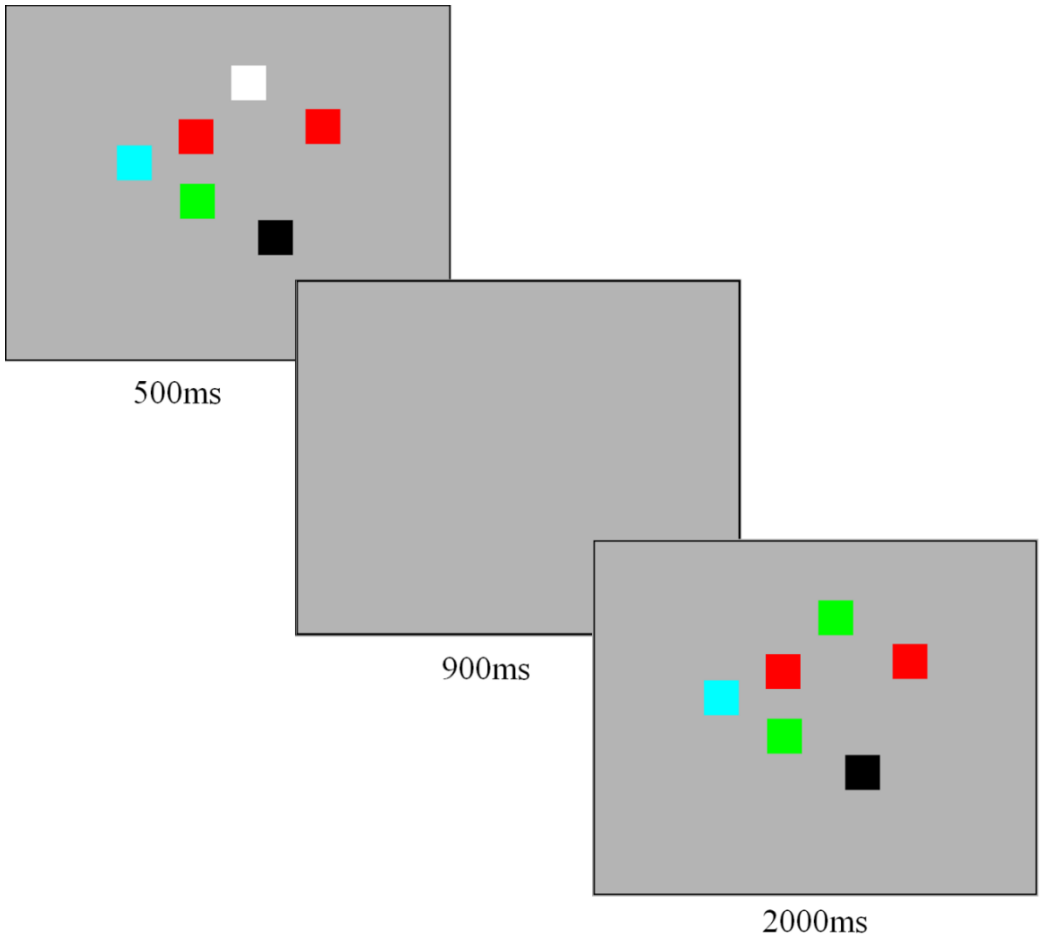


Figure 4-7. Example of a VWM test trial. Memory array stayed on screen for 500ms. Then after a 900ms retention interval, test array appear on screen for 2000ms waiting for user response.

4.5.3 Data analysis

Calculating VWM capacity

I calculated each participant's estimated VWM capacity using Cowan's formula [4]: If a participant can hold K objects in VWM from an array of S items, then on K/S trials the changed item should be one of those being held in VWM. This subject should be able to detect a change on K/S trials in which an item changed. This formula also corrects for guessing. Overall this formula is $K = S \times (H - F)$, where K is the VWM capacity, S is the size of the array, H is the observed hit rate and F is the false alarm rate.

Data Preprocessing

In experiment 2, I first calculated each of the 31 participants' solution optimality for each of the 20 TSPs. Solution optimality is defined as participant's solution length over the optimal solution length. For each of the 20 TSPs, I then calculated each participant's optimality percentile among all the participants. Then I take the median of this percentile among the 20 TSPs for each participant and excluded those participants whose median is over the 65th percentile. The intuition is that I want to exclude participants who perform much worse than the average for a majority of the problems.

Eye-tracking Data Analysis

Eye fixation data were exported from the Eye-tracker's Clearview 2.7 software with a filter setting of 100ms and 30 pixels, as recommended for mixed contents by the manufacturer. Fixation distances (in pixels) were then calculated according to our fixation distance definition (Figure 11). Those fixation distances were then distributed into 43 bins with centers on 10, 25, 40, 55… 640. According to the manufacturer's manual, the Tobii 1750 has an accuracy of 0.5° to

1.0°. In our experiment setting, 25 pixels roughly equals to 1 degree of visual angle. The count of the first bin centered at 10 pixels was discarded because it is well below the accuracy of the eye-tracker and is potentially noise. The counts of the rest of the bins and their corresponding fixation distances were fitted with three types of curve: Quadratic curve $p_1x^2+p_2x+p_3$, Exponential curve ae^{-bx}, and Power curve ax^{-b}. R^2 fitness was calculated for each type of curve fitting. To test if a curve is fitted significantly better by a power curve than an exponential curve, I used a Fisher's Z_r transformation on the Pearson correlation coefficients [65]. I define a curve's type to be exponential if R^2 fitness for the exponential curve is larger than that of the power curve and else verse. I use the exponential factor of the best fitted exponential curve as the decay rate unless the power curve fitness is significantly better than that of the exponential curve. R^2 fitness for quadratic curve was significantly worse than both exponential and power curve fitness.

4.5.4 Results

Our prior research has shown that even without global information available during the problem solving procedure, humans are able to get decent solutions for TSP but significantly worse than the condition when global information is visually available [48]. To eliminate the noise from individuals who make their decisions based only on local information or use a different task strategy than the one described in our model, I only analyzed the eye-movement data of the 22 subjects in the group with high overall solution optimality. A histogram of fixation distances was plotted for each subject. Then I fit three types of curves (quadratic, exponential, and power) to each subject's histogram of fixation distances. All of them fit very well to either an exponential

curve ae^{-bx} or a power curve ax^{b} (Rs > 0.97). Figure 13B shows three examples of curve fits to the fixation distances frequencies of three participants with different VWM size. Exactly as our model predicted, histograms generated from subjects with lower VWM capacity tend to be exponential while individuals with a higher VWM capacity tend to produce a power decay fixation pattern. Whether an individual has an exponential or power eye-movement pattern correlates very well with the individual's VWM size (R = 0.61, p < 0.003, N=22), as predicted by the model (Figure 4-8).

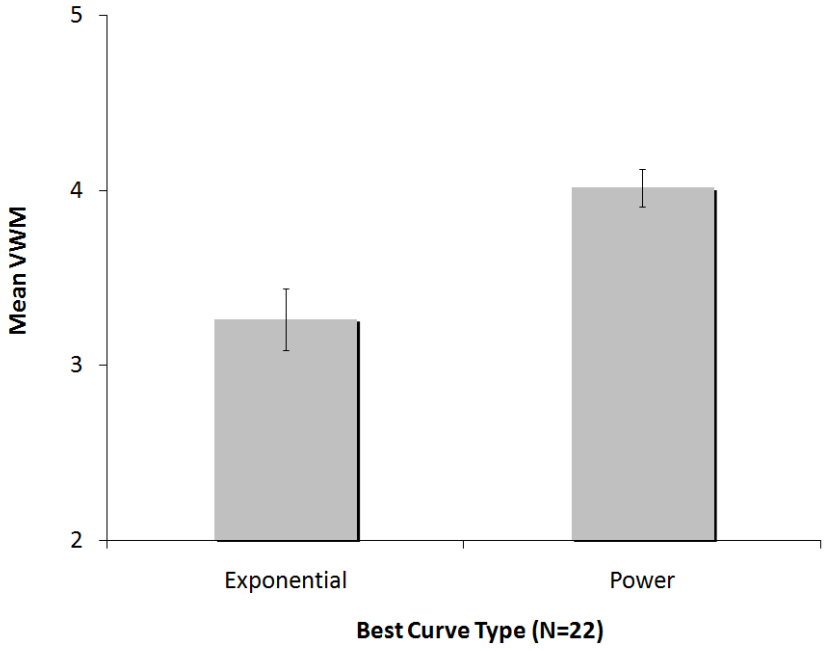


Figure 4-8. Histograms of fixation distances fit better to exponential curves than power curves for individuals with smaller VWM size and fit better to power curves for individuals with larger VWM.

There were sufficient exponential participants to further examine individual differences. For those

For those individuals with exponential eye-movement patterns, I found that the decay rate *b* of the

the exponential curves ae^{-bx} also correlates very well with the individual's VWM size (R = 0.8, p<0.001, p<0.001, N=15). (Figure 4-9)

Figure 4-10 shows histogram curves of three typical participants with different estimated visual working memory capacity.

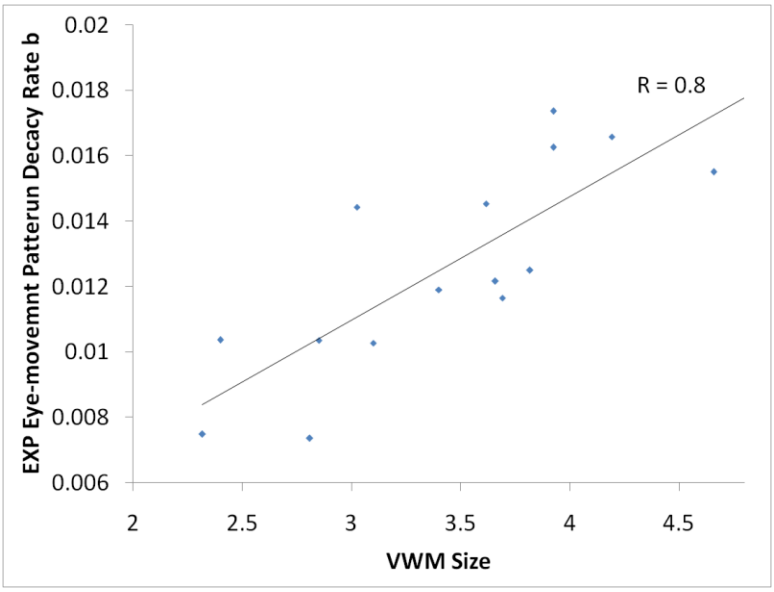


Figure 4-9. Correlation between VWM capacity and decay rate for exponential participants.

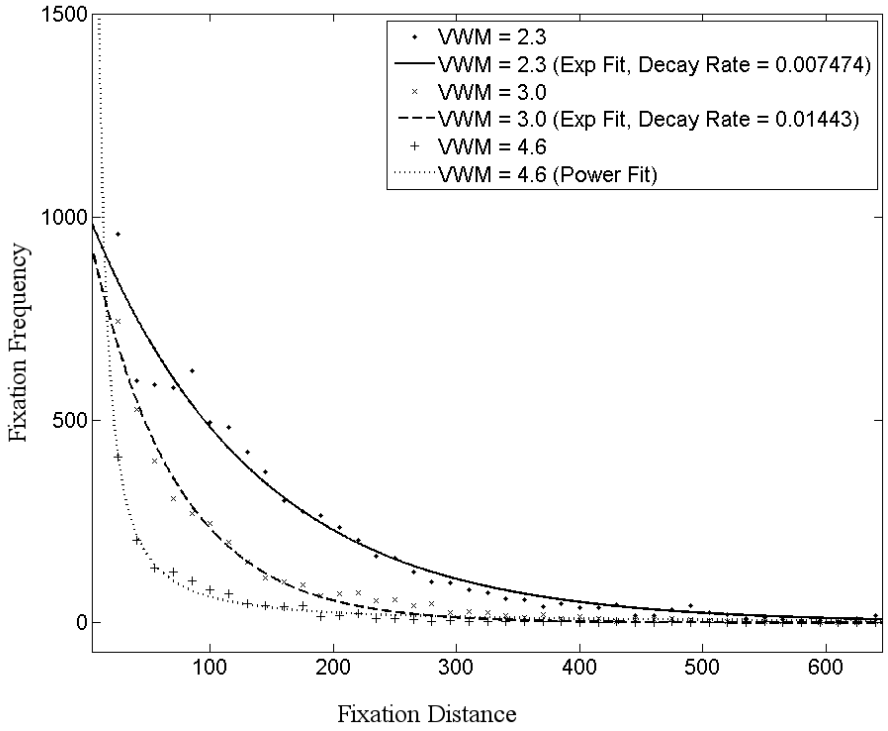


Figure 4-10. Examples of fixation distance patterns for three participants with different VWM capacities.

I also tested the effect of VWM size on TSP solution optimality. No significant correlation was found between individuals' VWM size and their solution optimality on any of the 20 TSP problems for all the 31 participants ($Rs<.29$, $p>0.12$). So the effect of VWM size on eye-movement pattern is not caused by trimming poorly performing participants. These results provide strong support for our hypothesis, since the relationship of individual VWM capacity and eye-movement pattern is exactly as predicted by our model.

I also compared the solution performance of our model with human data and existing models of human TSP solving. I calculated Pearson correlation on the means optimality of the 20 TSPs between 31 participants in the second experiment and those generated by the existing models. Compared to these other models, our model's solution for each problem correlates well

with human average solution (Our Model: R = 0.73; Convex Hull [66]: R = 0.63; Sequential Convex Hull [50]: R = 0.75; Nearest Neighbor [67]: R= 0.59; Pyramid [9]: R = 0.6; Kmeans [48]: R = 0.64), which further suggests that our model also captures problem solving as well as eye-movements.

4.5.5 Discussion

For over a decade, a lot of effort had been made to investigating the relationship between VWM and eye-movement [32, 68-71]. Our eye-tracking experiment results for the first time demonstrate a strong correlation between individual VWM capacity and eye-movement pattern in a non-trivial problem-solving scenario. Our model successfully predicated this result by an interaction between global perception, VWM, and visual attention in human complex problem solving. Although VWM capacity is extremely limited, the human visual system represents global information of different granularity in VWM where only the most relevant information is represented at the finest details. In this way, several chunks of information not only are able to capture the local details but also the big picture. During human complex visual problem solving, the contents of VWM constantly changes. The human visual system attends to local details when refining the global information into local ones and puts them into VWM. On the other hand, it also re-attends to global information when global information is needed but not represented in VWM. Individual VWM capacity plays a central role in deciding the precise pattern of visual attention. Overall, to seek an explanation of the long-standing mystery regarding how such a limited VWM could support such complex human visual reasoning abilities, our results provide strong evidence that the human visual system dynamically allocates VWM resources to represent

different granularity of global information and constantly updates its contents via attentional fixations. Although at any given moment the contents in human VWM are no more than several chunks of information, through this mechanism very complex visual information can be represented in a temporal manner to support highly effective human complex visual reasoning.

4.6 MODEL DETAILS

Step 1. Initialization

The current working set includes all points. The current point is set to be the starting point.

Step 2. Global Perception

Points in the current working set are grouped into M clusters using the K-Means clustering algorithm (MacQueen, 1967), where M is set to N (VWM size) in the first iteration, and smallest integer greater or equal to $\sqrt{N}$ afterwards. The K-Means Clustering Algorithm clusters N data points into M disjoint subsets Sj containing Nj data points so as to minimize the sum of squares criterion:

$$J = \sum_{j=1}^{K} \sum_{n \in S_j} | x_n - \mu_j |^2$$

where x_n is a vector representing the nth point and μ_j is the geometric centroid of the points in Sj. All the centroids are added into the collection of reference points in VWM, which was passed from the previous iteration.

The model then uses a spline-curve to connect the current point and all the reference points to sketch a path in a rough scale. The spline-curve is hypothesized to be a general smooth route through the centroids, which captures a general tendency of a globally sketched path.

Step 3. Identify current cluster and refine local information

All the points in the current working cluster are projected to its nearest points on the spline curve. If the number of points projected onto the part of the spline curve between the current point and the first reference point is more than 2, let the current working set to be this set of points and go back to step 2 for the next iteration. If it is not more than 2, continue to step 4.

Step 4. Move and rehearse global information

If the number of points projected on the spline curve between the current point and the next reference point is less than two, connect the current point to those points according to the sequence they projected onto the spline curve. Make the current working set to be the points projected onto the part of spline between the first and the second reference points. Discard the first reference point from the VWM.

If the number of reference points in the VWM is equal or less than 2, regroup unvisited points at the most global level and encode those centroids into the VWM.

Repeat this procedure until the number of unvisited points is less than the size of the VWM. Then find the shortest path for the rest of the points.

5.0 TOWARD A THEORY OF UNIFIED VISUAL WORKING MEMORY CAPACITY FOR GLOBAL AND LOCAL INFORMATION

How can such a limited visual working memory capacity support the complex visual problem solving ability that human cognition displays? In our attempts to investigate this big question, I used a cognitive modeling approach to test our hypothesis as described in the last chapter. Our experiment results were consistent with a number of principles of visual working memory that together with our model predicted eye-movement patterns during Traveling Salesman Problem solving.

One fundamental principle of visual working memory involves capacity limitations for global and local information. Here local information refers to basic component elements in a visual scene. Global information refers to summary or ensemble visual information which is often constructed from the basic visual elements. For example, in the Traveling Salesman Problem used in the last two chapters, each individual point is a piece of local information; the centroid for a group of points is a piece of global information. In an analogy to computer graphics, a piece of local information can be thought of as a single pixel and a piece of global information is the relationship among pixels. However, unlike with computers, human cognition has the ability to encode global information without necessarily encoding every bit of the included local information [1]. Do global and local information share the same capacity

limitation in visual working memory? Our TSP model and eye-tracking experiment suggest that the capacity limitation is unified across global and local information, although in a somewhat indirect fashion.

In this chapter, I directly investigate the hypothesis that both global information (summary information such as location of group centroid) and local information (such as exact location of a point) are stored in a unified visual working memory capacity. One chunk of local information, in this case the position of one point, occupies the same amount of visual working memory as one chunk of global (summary) information, in this case the centroid position of a group of points. Alvarez and Oliva [1] suggest that encoding a piece of global information requires a reduced amount of attention than encoding each individual pieces of local information in it. As an extension to this theory, I suggest that encoding information, global or local, has clear capacity limits, and further that the capacity limitation is unified across global and local information rather than separate capacities for each. As a strong test of our hypothesis, I adapt the paradigm used by Alvarez and Oliva.

5.1 EXPERIMENT FOUR

5.1.1 Methods

5.1.1.1 Equipment

The experiment was displayed using the Psychophysics Toolbox for Matlab [72, 73] on a 17" LCD screen with an embedded eye-tracker Tobii 1750.

Participants

41 undergraduate students in the University of Pittsburgh participated in the experiment to fulfill course requirement.

5.1.1.2 Materials

Each participant finished 120 trials in a global/local tracking session, followed by a VWM test session.

5.1.1.3 Procedure

Tracking and missing-item localization task. In the first session, participants performed a global/local multiple-object-tracking task followed by a missing item localization task. Global/local multiple-object-tracking task was first used by Alvarez and Oliva [1]. I adapted this experiment paradigm with some simple extensions. In each trial of the experiment, two types of targets are displayed on the screen: local targets and global targets. Each local target is a single black dot. Each global target is a group of four black dots. As shown in Figure 5-1, the screen is divided into five different regions: top, bottom, left, right and center. Local targets move randomly within the center region. Points in each global target move randomly within the left, right, or top regions. Once a trial starts, targets are displayed within the selected regions of the screen. Two red diagonal lines were displayed on the screen during each trial. When each trial starts, local targets flash off and on for 2 seconds. Then, all points move for a random duration between 6 to 10 seconds before stopping again.

Participants were instructed to perform two tasks simultaneously in each trial. The primary task is to count the total number of times that local targets cross either of the diagonal

lines. In addition, when the points stop moving, one global target (all points in that global target) disappeared from the screen. The secondary task is to locate the centroid of the four points in that global target after it disappeared from the screen. Participants were instructed to first use the mouse to click on the centroid of the missing group of points (secondary, global targets task) then type in the number of times that local targets went across the diagonal lines (primary, local targets task). After each trial, feedback was provided on the performance of both global and local tasks. An overall score is also displayed as the feedback on the overall performance in each trial. Performance on the local task has a two-thirds weight on the overall score and performance on global task has a one-third weight.

Conditions. The number of local targets and the number of global targets were manipulated within-subjects as experiment conditions. However, it was not necessary to investigate all possible combinations, as I was primarily interested in main effects rather than interactions. Thus, I manipulated the number of local targets, while holding global targets constant at 2, and then manipulated the number of global targets, holding the number of local targets constant at 4 (see Figure 5-2). These values were selected to move participants from just around capacity limitations to just beyond capacity limitations. There were 20 trials per cell (120 trials total), with trial type randomly generated on each trial.

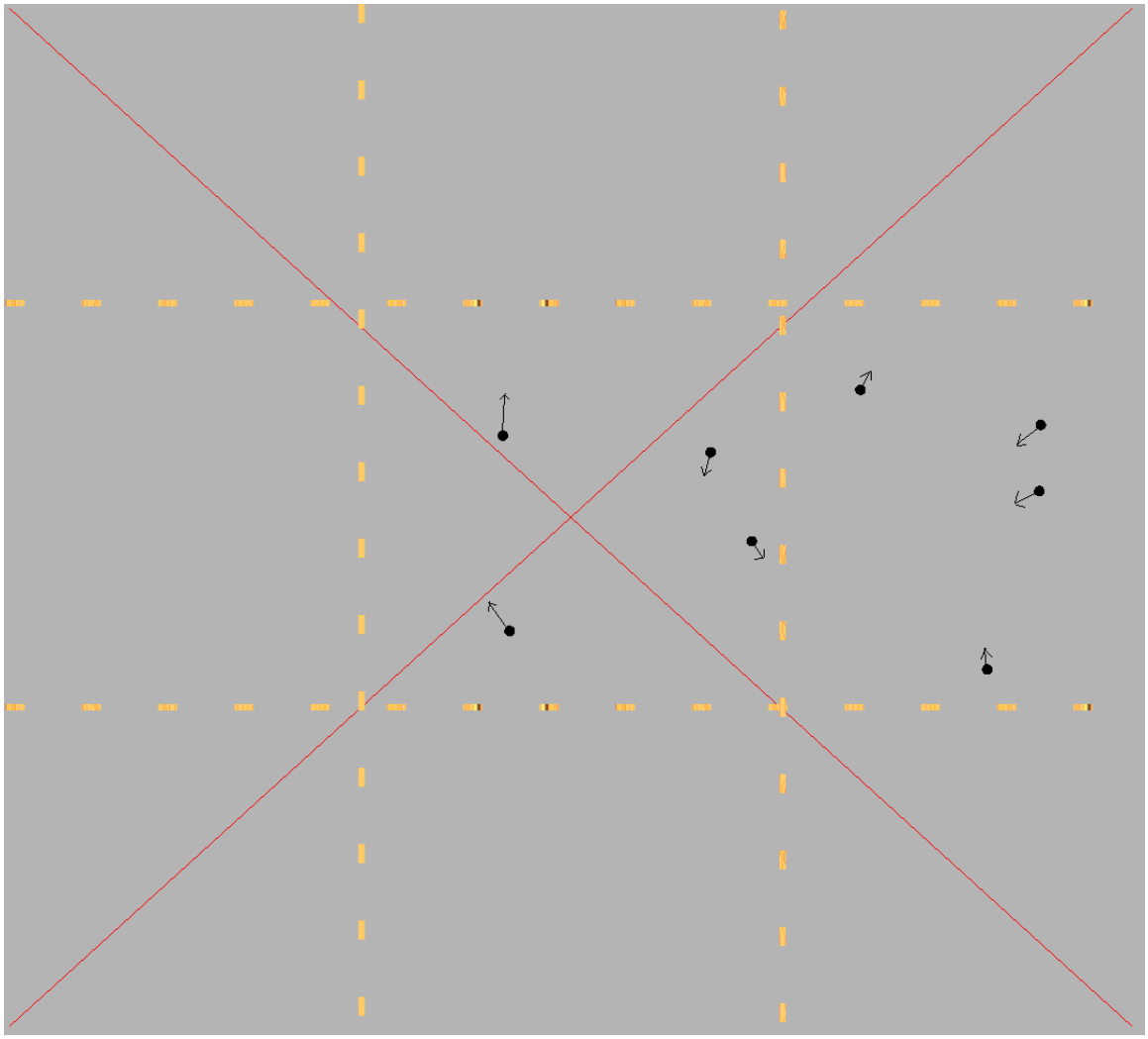


Figure 5-1. Example of an experiment trial (one distracter condition): groups of points randomly move within their own region. (Yellow dashed lines are invisible in the actual experiment.)

		Number of Global Targets		
		1	2	3
Number	3			
of	4			
Local	5			
Targets	6			

Figure 5-2: Experiment conditions are illustrated as the highlighted areas

Performance metrics. To measure the performance of participants on each experiment trial, error was defined in terms of the following two metrics:

1. Local Error is defined as the percentage of offset between the response on the counting task and correct answer (e.g., reporting 7 rather than 8 is (7-8)/8 x 100 = 12.5%).
2. Global Error is defined as the Euclidean distance (pixels) between the response mouse click and the actual location of the global target centroid.

Furthermore, participants' eye-movements were recorded during this session of the experiment. I defined the following two metrics on eye-movements:

1. Percentage of Center Fixations: the percentage of fixations within the center region of the screen in which the local targets move around.
2. Time since Last Look: the number of milliseconds between when points stopped moving and the last fixation in the global region being tested.

Visual working memory task. In a second session, I estimated the visual working memory capacity of each individual participant using the same procedure as used in Experiment 3.

5.1.1.4 Predictions

The experiment is designed to test the capacity limitation of visual working memory for global information and local information. We can think of visual working memory as a backpack, with global and local information as food and water that are contained within the backpack. There are several hypotheses regarding how they occupy the backpack.

The shared capacity hypothesis (i.e., our main hypothesis) states that global and local information occupy the same visual working memory space. This hypothesis suggests that global and local information are like canned food and bottled water. When global and local information are loaded in visual working memory, they share the same space limitations.

The separate capacity hypothesis states that global and local information occupy different visual working memory spaces. This hypothesis suggests that the backpack has two different compartments, a food compartment for holding food and a hydration compartment for holding water. One cannot load food into the hydration compartment nor put water into the food compartment.

Within the shared capacity hypothesis, there are two different hypotheses for the space that each chunk of global or local information occupies in visual working memory: 1) A unified

capacity hypothesis says that either global or local information occupies the same amount of space in visual working memory in a shared space. This hypothesis suggests that each food can is of the same size as a water bottle. 2) A non-unified capacity hypothesis says that global and local information share the same capacity limitation in visual working memory but they occupy a different amount of space for a single chunk of information. This hypothesis suggests that either food cans are bigger than water bottles or water bottles are bigger than food cans, but they share the same space in the backpack.

Between the shared capacity hypothesis and the separate capacity hypothesis, another possible way for global and local information to occupy the visual working memory is the single-directional share hypothesis. Single-directional share hypothesis says that global and local information occupies different spaces in the visual working memory with the exception that a one directional share is possible. For example, one direction is that when the space for global information is full, global information can be stored in the space of local information (but not vice versa). Another direction is the other way around (i.e., local information in the global space but not vice versa). In our backpack analogy, this hypothesis suggests that the hydration compartment would only be able to hold water, but the food compartment, although are primarily used to hold food, can also be used to hold bottle water if the hydration compartment is full.

If the separate capacity hypothesis is true, conditions that vary the number of local targets would not affect the performance of the global task as measured by global error; similarly, conditions that vary the number of global targets would not affect the performance of local task as measured by local error. In other words, since the separate capacity hypothesis suggests that food and water are loaded into different compartments of the backpack, whether or not one

compartment is full should not affect its ability to hold another type of element in the other compartment.

If the single-directional share hypothesis is true, then when hydration compartment of the backpack is full, bottled water can be loaded into the food compartment, but food can't be loaded into the hydration compartment even if the food compartment is full. Therefore, either of the following effects should be observed depending on the direction of share, but not both: the number of local targets affects global error, or the number of global targets affects local error.

If the shared capacity hypothesis is true, then the number of global targets should affect local error *and* the number of local targets should affect global error as well. In our backpack analogy, since canned food and bottle water shares the same space in the backpack, when the backpack is full there is no more room for either of them.

However, if the non-unified capacity hypothesis is true, which means that canned food and bottled water are of different sizes, then when the backpack is almost full, it may be able to fit another bottled water (or canned food) but not another canned food (or bottled water). So if this hypothesis is true, when the visual working memory is loaded close to its capacity limitation, adding another local (or global) target may not affect the performance but adding another global (or local) target may cause a decrease of performance measured by both global and local error.

If the unified capacity hypothesis is true, when the capacity limit is reached, adding either a global or local target will cause a decrease in performance measured by both global and local error. In other words, when the canned-food and bottle-water are of the same size, if the backpack can't fit another canned-food, then it can't fit another bottle-water, and vice versa.

The basis for all the hypotheses described above is that the changes of both global and local error are caused by the capacity limitations of visual working memory. So before the

capacity limitation is reached, the change of conditions, in number of global or local targets, should not affect the performance of the corresponding global or local task. After the capacity limitation is reached, decay in performance should be observed in local (global) task when the number of local (global) targets increases.

5.1.2 Results and discussion

Forty-one participants each completed 120 trials of the global/local multi-object tracking task. Out of the total 4920 trials, I trimmed approximately 3% of the trials in which participant's performance was much worse than the rest—in those trials, either global or local error was higher than the threshold of 200 pixels for global error or a 10 counting offset for local error. I assume that in these high error trials, participants either had a large lapse of attention or misclicked/mistyped their answers.

5.1.2.1 Condition effects on global and local error data

When the number of global targets is fixed to two, Figure 5-1 illustrates the effect of the number of local targets on local error. A repeated measures ANOVA revealed a significant effect for the number of local targets on local error ($F(3,38)=87.5$, p <.001). A LSD pair-wise comparison revealed significant differences when the number of local targets increased from 4 to 5 (M = -.076, SEM = .009, p < .001) and from 5 to 6 (M = -.049, SEM = .007, p < .001). But when the number of local targets increased from 3 to 4, the effect on local error was not significant (M = -.007, SEM =.021, p =.73). This pattern is consistent with a capacity threshold being crossed.

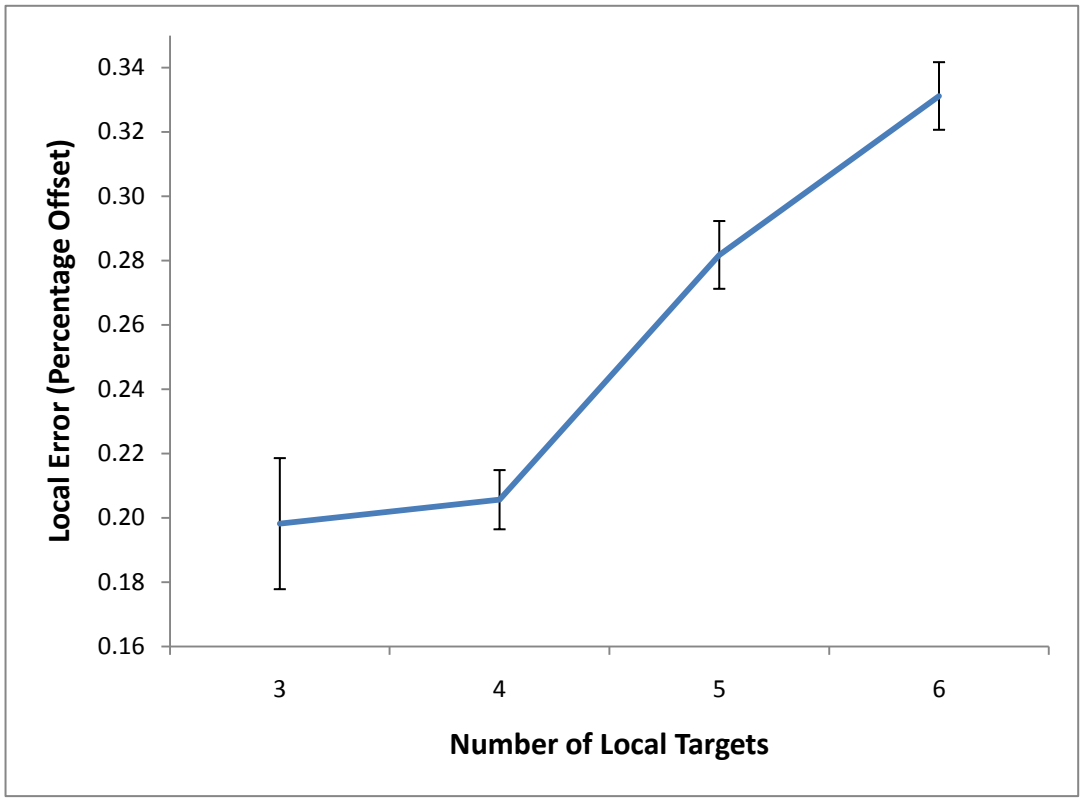


Figure 5-1: Effect of number of local targets on local error, with standard error bars.

When the number of local targets is fixed at four, Figure 5-2 illustrates the effect of the number of global targets on global error. A repeated measures ANOVA revealed a significant effect for the number of global targets on global error ($F(2,39)=5.74$, $p = .007$). A LSD pair-wise comparison revealed a significant effect when the number of global targets increased from 2 to 3 (M = -7.0, SEM = 2.1, $p = .002$). But the difference from 1 to 2 (M = 2.8, SEM = 2.31, $p = .23$) and 1 to 3 (M = -4.2, SEM = 2.7, $p = .14$) was not significant. Given that a curvilinear trend is highly unlikely and the near significance of the 1 vs. 3 comparison, I interpret these results as $1 = 2 < 3$, which again is consistent with a capacity threshold being crossed.

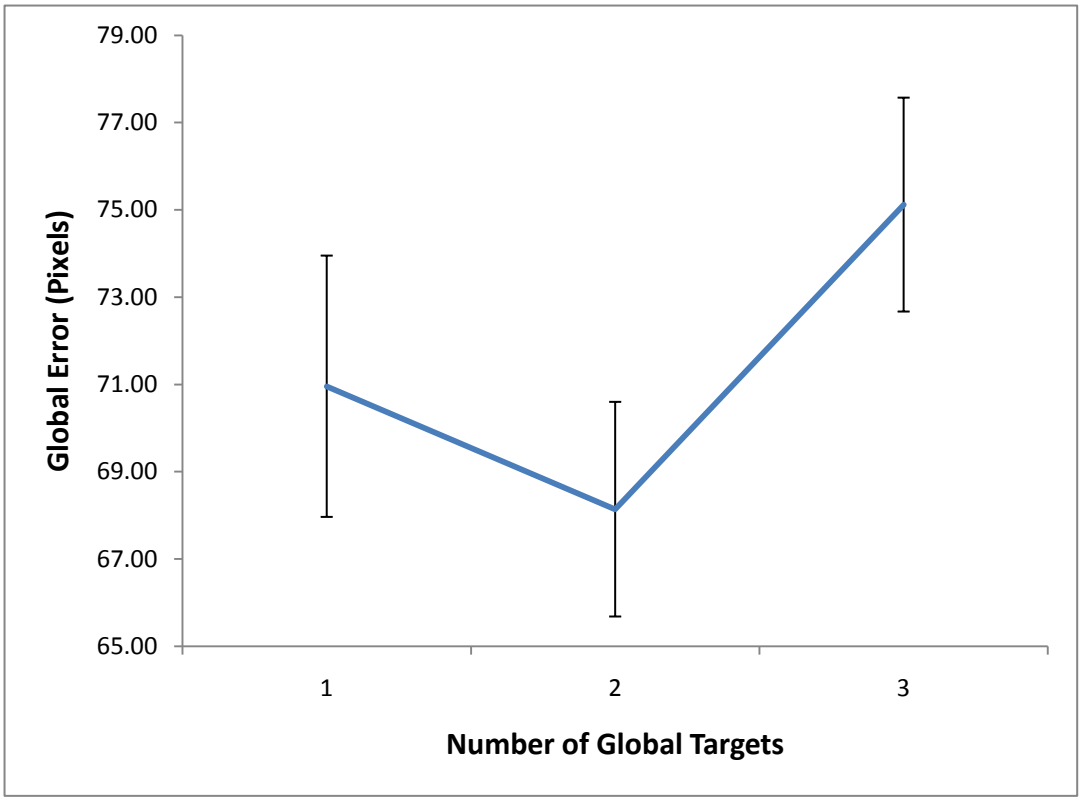


Figure 5-2: Effect of number of global targets on global error, with standard error bars.

When the number of global targets is fixed to two, Figure 5-3 illustrate the effect of the number of local targets on global error. A repeated measures ANOVA revealed a significant effect for the number of local targets on global error ($F(3,38) = 7.23$, $p < .001$). A LSD pair-wise comparison revealed significant differences when the number of local targets increased from 4 to 5 (M = -5.3, SEM = 1.9, p=.009) and from 5 to 6 (M = -3.7, SEM = 1.68, p =.035). But when the number of local targets increased from 3 to 4, the effect on global error was not significant (M = .58, SEM = 1.6, p =.72). Again, this pattern is consistent with a capacity threshold being crossed, interestingly at the same point (i.e., between 3 and 4) for global and local error as a function of the number of local targets.

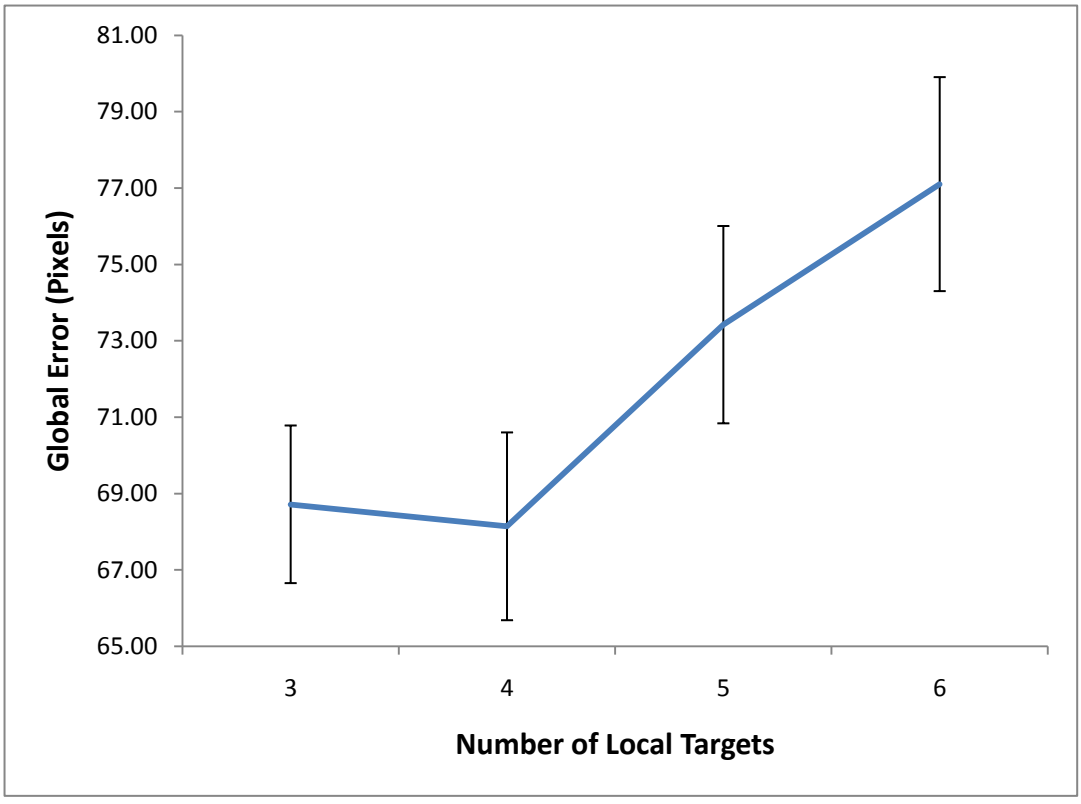


Figure 5-3: Effect of number of local targets on global error, with standard error bars

When the number of local targets is fixed at four, Figure 5-2 illustrates the effect of the number of global targets on local error. A repeated measures ANOVA revealed a significant effect for the number of global targets on local error ($F(2,39) = 7.4$, $p = .002$). A LSD pair-wise comparison revealed a significant effect when the number of global targets increased from 2 to 3 (M = -.032, SEM = .008, p =.007). But the difference from 1 to 2 (M = -.007, SEM = .008, p = .83) was not significant. Again the pattern is consistent with a capacity threshold being crossed and again the capacity breakpoint (after 2 global targets) is the same for both global and local error measures.

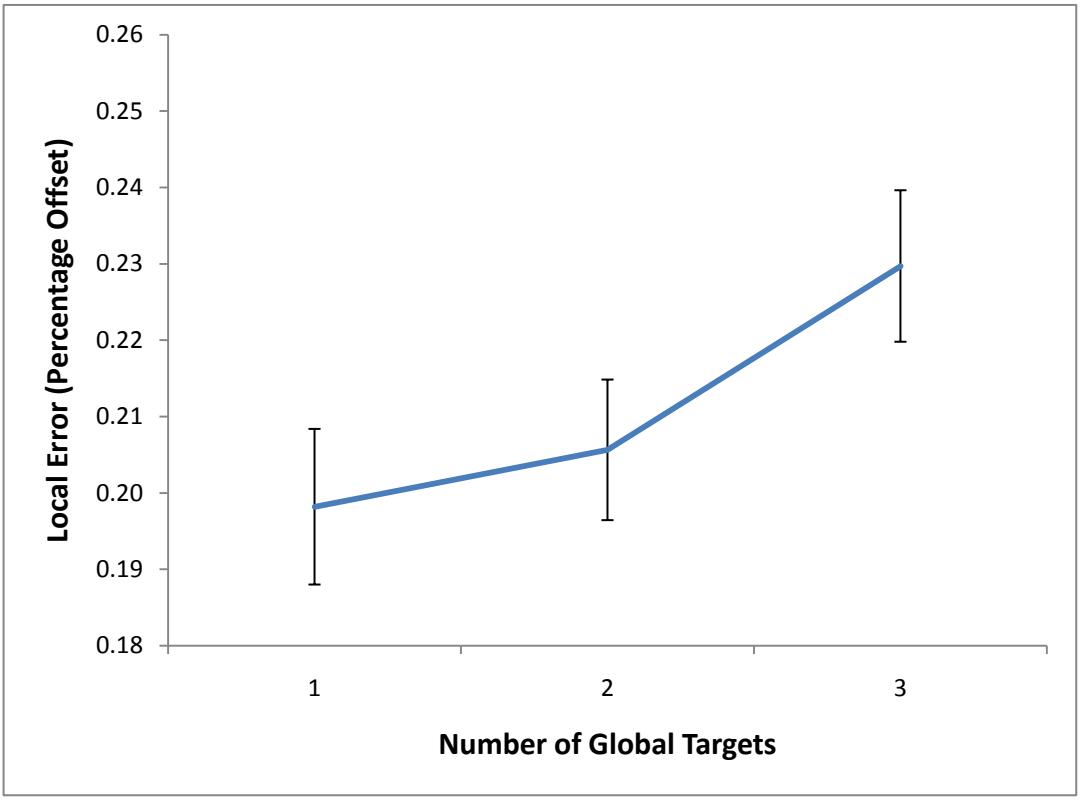


Figure 5-4: Effect of number of global targets on local error, with standard error bars

5.1.2.2 Correlations with visual working memory capacity

Visual working memory capacity for each participant was estimated using the same procedure as in Experiment 3. Among the forty-one participants, two participants did not completely finish session two of the experiment and were excluded from the analyses involving the visual working memory capacity data.

As shown in Figure 5-5, participants' estimated visual working memory capacities significantly correlated with the mean local error generated by each subjects (n=39, R = -0.47, *p* = .001).

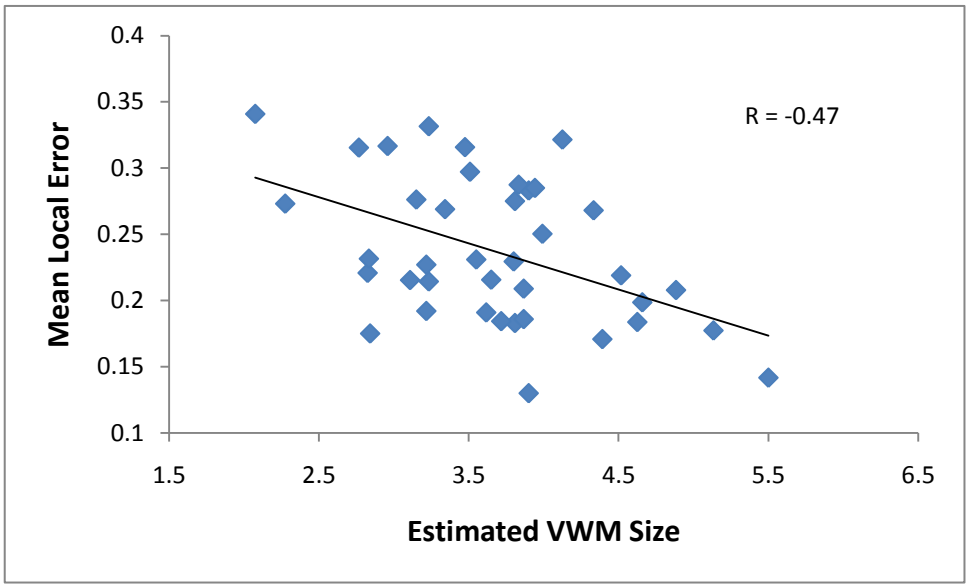


Figure 5-5: Estimated visual working memory size vs. mean local error

However, participants' estimated visual working memory capacity did not correlate significantly with the mean global error generated by each subjects (n = 39, R = 0.08, *p* = .31). Participants' mean global error did not correlate with mean local error either (n = 39, R = 0.05, *p* = .75), possibly suggesting some kind of differential strategic tradeoff across subjects for how they

weighted the primary and secondary tasks, which would interfere with the analysis of working memory capacity effects on error within a given measure.

5.1.2.3 Eye-Movement results

To further verify that this strategic tradeoff is employed by participants, we examined their eye-movement patterns. As illustrated in Figure 5-6, the number of local targets had a significant effect on the percentage of center fixations ($F(3,38) = 10.7$, $p < .001$). A LSD pair-wise comparison revealed significant difference when number of local targets increased from 4 to 5 (M = -.006, SEM = .003, $p = .03$) and 5 to 6 (M = -.005, SEM = .002, $p = .02$), but when number of local targets increased from 3 to 4 the effect on percentage of center fixations is not significant (M = -.003, SEM = .004, $p = .35$). However, note that effect overall is small and the means across conditions are all very close to 1, suggesting that participants did treat the primary task as the primary task and kept their eyes on the center region almost exclusively.

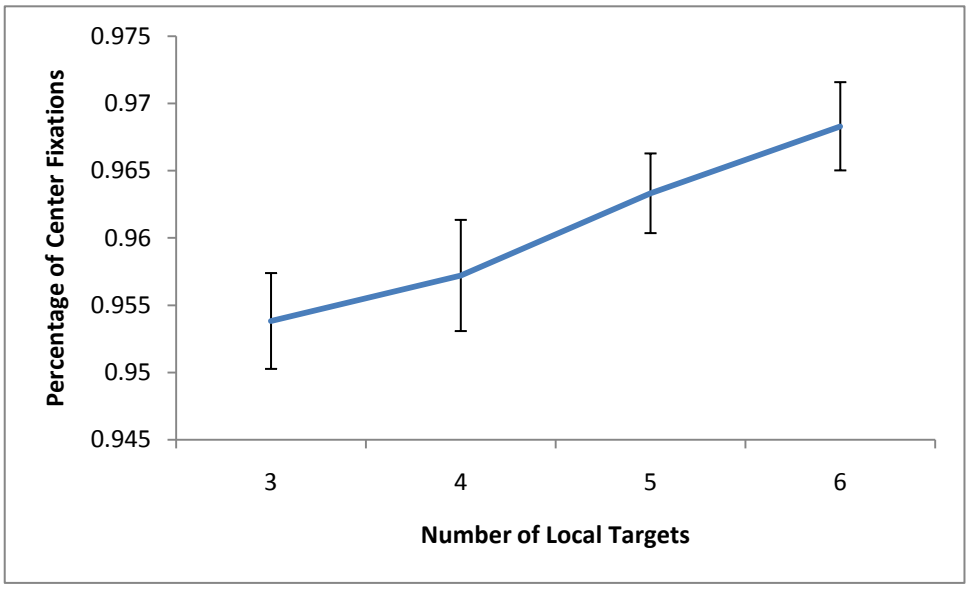


Figure 5-6: Effect of number of local targets on percentage of center fixations, with standard error bars

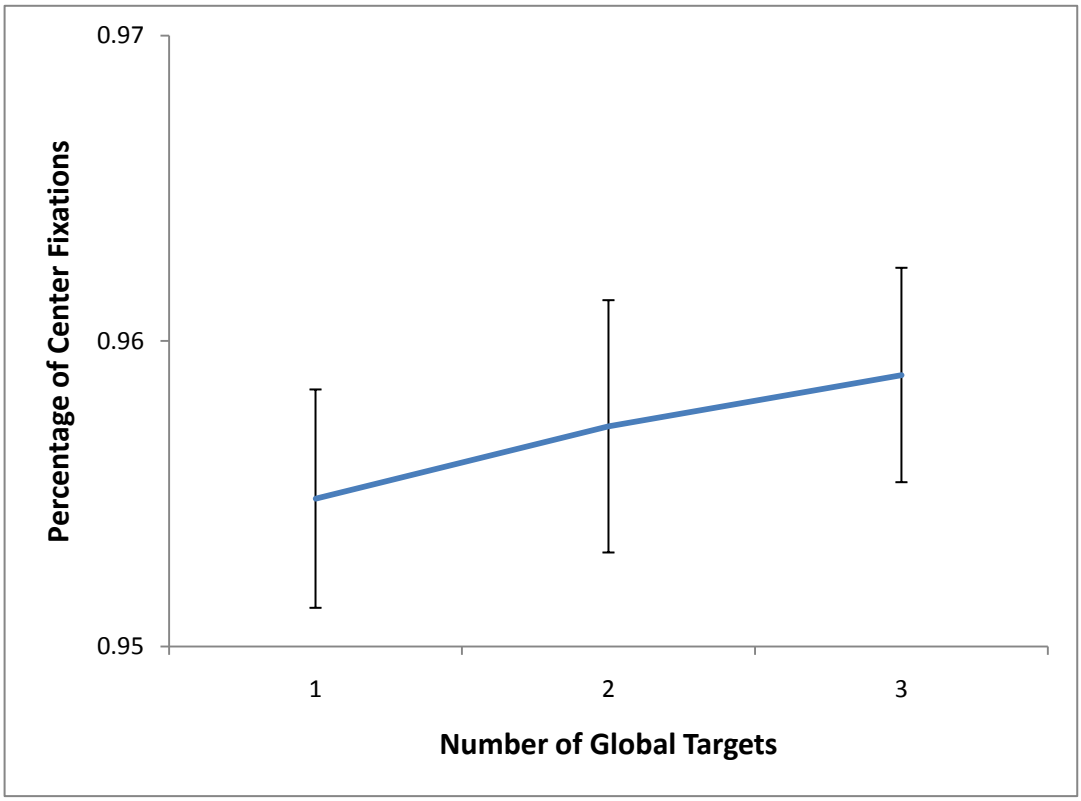


Figure 5-7: Effect of number of global targets on percentage of center fixations, with standard error bars

Figure 5-7 shows the average percentage of center fixations for each condition on number of global targets. Repeated measures ANOVA revealed no significant effect of the number of global targets on the percentage of center fixations ($F(2,38) = 1.69$, $p = .20$). Thus, having more global targets did not lead to move fixations away from the center region, consistent with the global tracking task being the secondary task.

Time since last global fixation did not change as a function of the number of local targets ($F(3,36) = 1.7$, $p = .33$), as illustrated in Figure 5-8, ruling out the possible cause of the effect of number of local targets on global error as being due to a decay function (i.e., time since last viewed).

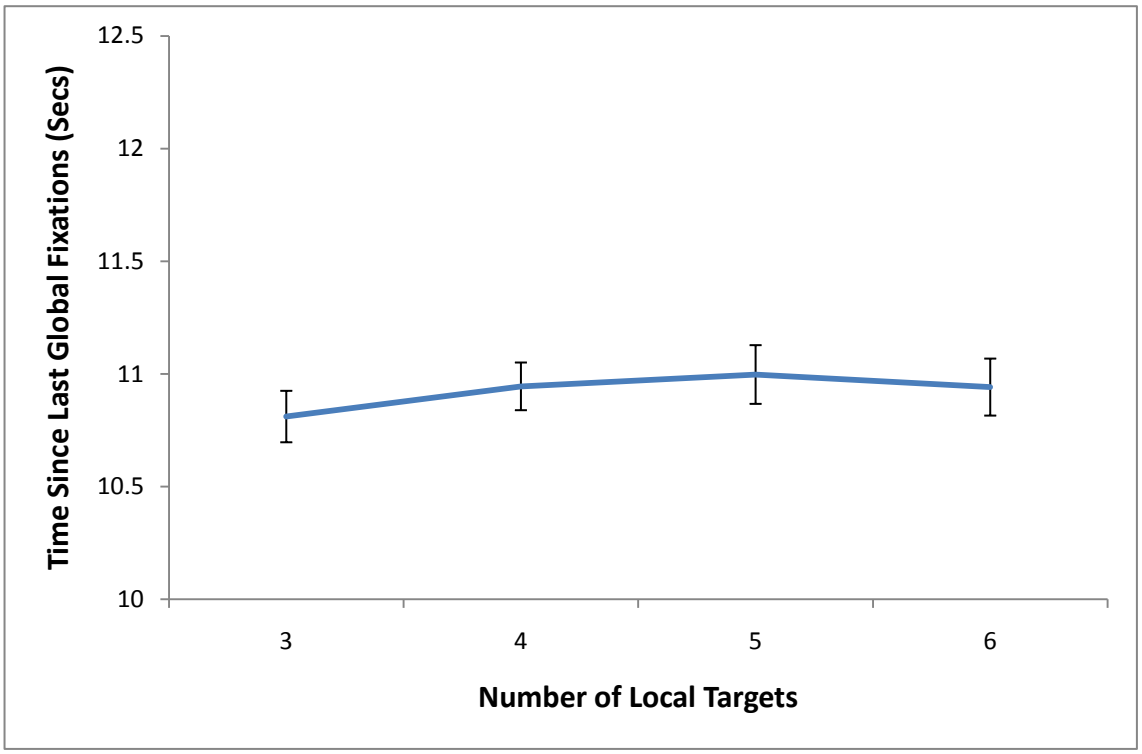


Figure 5-8: Effect of number of local targets on time since global fixation, with standard error bars

The number of global targets has a significant effect on time since last global fixation ($F(2,37) = 29.0$, $P < .001$). As shown in Figure 5-9 and revealed in a LSD pair-wise comparison, time since last global fixation do not differ significantly when number of global targets increases from 1 to 2 (M = -.19, SEM = .11, $p = .09$), but did differ significantly when number of global targets increases from 2 to 3 (M = -.67, SEM = .09, $p < .001$). Thus, it remains possible from this data that the effect of global targets on global error was perhaps partially the result of a decay function (i.e., time since last fixated), although the time difference between end-points was smaller than 10% and thus not likely to be the primary source of the condition effect on error.

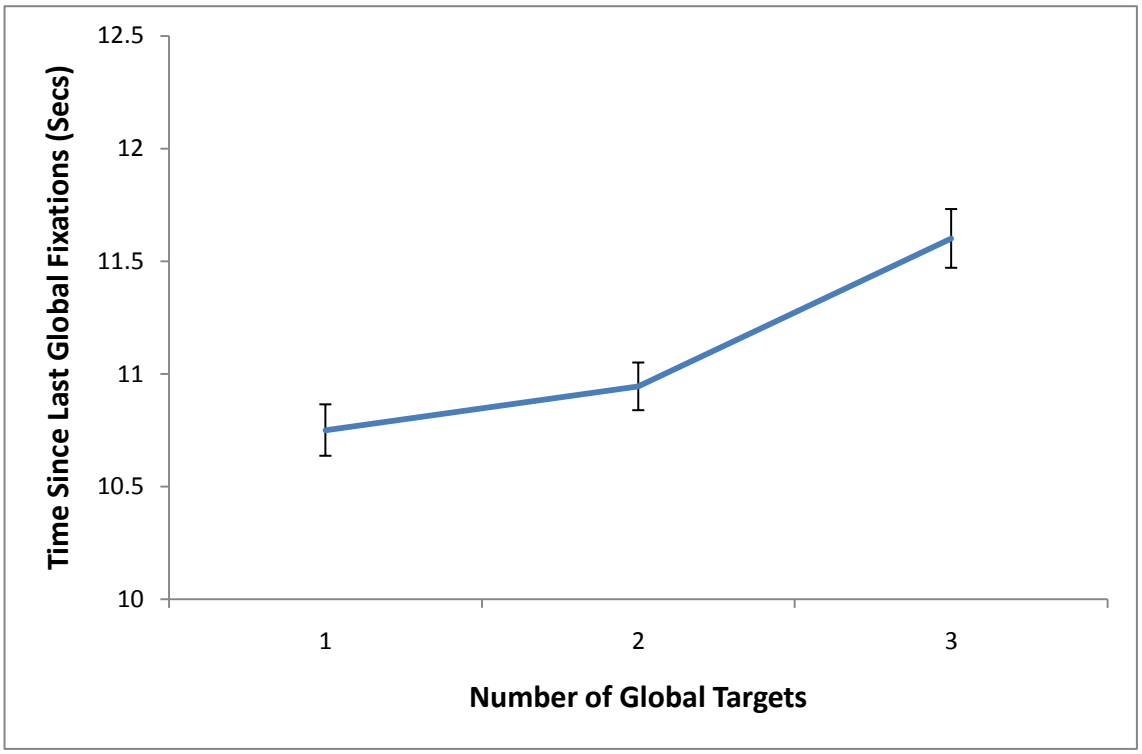


Figure 5-9: Effect of number of global targets on time since last global fixation

Overall time since last global fixation does not significantly correlate with global error (R(4483) = -.026, p = .08), further suggesting that global fixation timing are not a likely source of the effect of global targets on global error (i.e., that the effect is more likely one purely of capacity limitations).

5.1.3 General discussion

In this experiment, two sets of conditions were designed to vary the number of global and local targets. Two performance metrics, global and local error, were measured.

First, our experiment results show clear capacity threshold effects. The number of local targets does not affect local error until it is greater than 4 (when the number of global targets was fixed at 2). As expected, this result is consistent with the FINST theory [35, 36], which also posits a capacity limitation of 4 or 5 local elements that can be tracked. Similarly, the number of global targets does not affect global error until it is greater than 2 (number of local targets was fixed to 4). In our backpack analogy, performance are similar before the backpack is full; additional errors are generated when each additional element are added to a fully occupied backpack. So our results suggest a similar fixed capacity limitation exists for holding global information. Another possible explaination for the increase in global error with number of global targets is caused by the less recent fixations on global targets as shown in Figure 5-9. However, the correlation between global error and time since last global fixation is so low ($R(4483) = -.026$, $p = .08$) that this explanation is implausible.

Second, our experiment results also support the shared capacity hypothesis and therefore are against the separate capacity hypothesis. When the number of local targets increased from 4 to 5 and 5 to 6, global error significantly increases (Figure 5-3). If local information and global information had separate capacity limitations, increases in the number of local targets would not cause increases in global error. Furthermore, this increase in global error was not found when number of local targets increased from 3 to 4 (Figure 5-3), where the same pattern was found in number of local targets on local error (Figure 5-1). In addition, the increase in global error with number of local targets was not caused by a changing percentage of center fixations alone (Figure 5-6), because there is no significant correlation between percentage of center fixations and global error.

From the other direction, the number of global targets also shares this pattern of change, where there is no effect on local error when increased from 1 to 2 and a significant effect when increased from 2 to 3 (Figure 5-4). So this sharing of capacity limits is not single-directional. This consistent pattern of effects between two sets of conditions on two performance metrics (Figure 5-1, Figure 5-2, Figure 5-3, Figure 5-4) provides evidence for the unified shared capacity hypothesis, since both global and local error started to increase when the total number of targets (local + global) changes from 6 to 7; and both types of error remained at the same level when the total number of targets changes from 5 to 6.

Similar to many dual task experiments, strategy difference were observed acrosss conditions. First, as the number of local targets increased, the percentage of center fixations increased as well (Figure 5-6). This increasing attention on the primary task is expected, since the participants were instructed to focus on primary task first and they received twice the rewards for their primary task (local task) performance as for their secondary task (global task) performance. The hypothesis is that the primary task has a higher priority in requesting attentional resources such as visual working memory and eye-movements.

Across participants, the correlations between estimated visual working memory capacity and global/local task performances also support memory capacity explanation of the effects of targets on performance. There is a significant negative correlation between estimated visual working memory capacity and the mean local error ($R(41) = -.46$, $p = .001$). So the local error decreased as visual working memory capacity increased, as shown in Figure 5-5. In other words, when the size of "backpack" increases, fewer "food cans" were left out due to insufficient capacity. However, there is no significant correlation between estimated visual working memory capacity and mean global error ($R(41) = 0.08$, $p = .31$), which suggest the individuals with higher

visual working memory capacities allocated more visual working memory to the local task but not to the global task.

5.1.4 Conclusion

In this experiment, I explored and tested several possible hypotheses regarding the capacity limitations on visual working memory for global and local visual information:

1. Separate capacity hypothesis: Global and location information are held in different compartments in visual working memory, with the global compartment only holding global information and the local compartment only holding local information.
2. Single-directional share hypothesis: Two compartments are there in the visual working memory: a special compartment and a generic compartment. For example, the local (global) compartment only holds local (or global) information and the generic compartment can hold both.
3. Shared capacity hypothesis: There is only a single compartment in the visual working memory. Both global and local information share the same space. Within this hypothesis there are two sub-hypotheses regarding how much space each piece of global and local information occupies the visual working memory:
    a. Non-unified shared capacity hypothesis: Global and local information share the same space in the visual working memory. The space occupied by each piece of global information is different from each piece of local information.

b. Unified shared capacity hypothesis: Global and local information share the same capacity in the visual working memory. Each piece of global information occupies the same amount of as each piece of local information.

Overall, our experiment results support the unified shared capacity hypothesis best. This hypothesis is also suggested by our previous Traveling-Salesman-Problem experiment and its corresponding model results.

5.1.5 Discussion

At the first glance, this experiment results and conclusions may seem contradictory to what Alveraz and Oliva suggested in [1] that global information is freely available to human visual system. However, by taking a closer look, it is not hard to find that this unified shared capacity hypothesis is complementary to what Alveraz and Oliva found and is consistent to their experiment results as well. Consistent with Alveraz and Oliva, I find that global information is freely available to human vision in the sense that individual piece of local information inside global information is not represented. In addition, the storage of global information in VWM is limited in a unified shared capacity together with local information. In the centroid test condition in [1], the trials are basically identical to the condition in our experiment with number of local targets set to 4 and number of global targets to 1. So there were in total 4+1=5 targets, if we think centroid of a group of points as one target. Participants were able to track those targets at lease better than chance, because the total number was not quite over the shared limitation. However, in their individual condition when exact location of a point was asked, the total number of targets is 4 + 4 = 8. In this case, since the total number is a lot more than what can be

hold in the limitation, and the participants were told to focus on the primary targets, the "distracters" were not enable to fit into the capacity limitation. Thus, in this condition, the performance to report a single distracter location was no better than chance.

The unified shared capacity hypothesis extended what Alveraz and Oliva found in [1]. Our visual system not only can represent global information without representing each individual local element, but also holds each piece of global information in the same unified capacity as each piece of local information.

6.0 CONCLUSIONS AND DISCUSSION

Even with only about 2-5 slots of visual working memory (VWM), humans are able to solve complex visual problems in real time. To explain the paradox between human VWM constraints and human visual complex problem solving ability, in this dissertation study I examined the role of global information processing in complex visual problem solving and the ways in which VWM resources are allocated to global and local information through eye-movements. In experiment one, I investigated how global information affected performance in human traveling salesman problem solving and how it is represented internally and externally. Based on the observations and results from this first experiment, I build my first TSP model and evaluated this model by comparing its task performance against the task performance humans, especially in comparison the fit of prior TSP models.

Experiment one showed that global information plays a major role in human complex visual problem solving. Without the presence of global information, participants performed significantly worse in the task. Furthermore, the results suggest that global information was not just perceived in the beginning and kept in working memory for the entire duration of problem solving, but rather there are interactions between the external representation and internal working memory at each phase of problem solving.

To further explore the mechanism underlying this interaction, I conducted the second experiment with eye-movement data recorded during problem solving. Histogram of fixation distance from the current path has shown to have a very regular decreasing pattern and was fitted very well by exponential curves. I extended the first TSP model by taking the VWM constraint into account and simulated the eye-movements by recording fixations whenever a piece of information (global or local) is encoded into VWM. The eye-movement pattern generated by this second model captured the aforementioned eye-movement pattern very well. In this second TSP model, the VWM size is an adjustable parameter. When this parameter changes, the decay rate of the resulting fixation-histogram also changes. An individual difference in this decay rate was also found in the experiment result. This observation supports our model and led to the hypothesis that individual differences in visual working memory caused the quantitative differences in eye-movement pattern. Results of the second experiment indirectly suggest that a very regular eye-movement pattern is a result of a VWM constraint. Since VWM can only hold a very limited amount of information, one often has to re-fixate to re-encode external representations into VWM. The second experiment left open the question that whether the individual differences in eye-movement pattern is caused by individual differences in VWM size. Our model suggested that this hypothesis is true in an indirect way.

To further directly test this hypothesis, I did the third experiment. In the third experiment, after the first session of TSP solving, a change-detection task was used to estimate the participant's VWM size. The third experiment result has shown a clear correlation between the individual's VWM size and the decay rate of fixation histogram when performance was within a certain range. Thus it appears that individual differences in VWM size were the cause of the difference in eye-movement pattern. Based on this observation and the observation that our

second TSP model also adequately fitted the performance data despite having a very reduced memory capacity in comparison the first model, the underlying interaction between internal VWM and external representation became clearer. Our model and experiment results suggest that global (summary) information and local information were kept in a unified VWM capacity. At each instance, a balanced amount of global information and local information were stored in VWM to support the reasoning process. Eye-movements were directed to encode global information and break global information into local information components. In this way, complex visual information can be represented in a temporal manner. It opens a possible explanation for the paradox I'm trying to address in this dissertation: Even though VWM capacity is so limited, eye-movements balance the ratio of local and global information in VWM, enabling attention to fine details without losing the big picture. This explanation relies on the unified share VWM hypothesis that my second TSP model and third experiment's results supported in an indirect way.

In the fourth experiment, I further directly tested this unified VWM capacity hypothesis by using a different experiment paradigm. The fourth experiment employed a multiple-object-tracking task and was designed to test multiple hypotheses regarding how global and local information is represented in VWM memory. This version of multiple-object-tracking task was first employed by Alveraz and Oliva [1]. Participants were asked to keep track of two kinds of moving objects, global and local, at the same time. Although the multiple-object-tracking task is a different experimental paradigm from traveling salesman problem solving, they share the same basic unit of local information, which is the position of a single point, as well as the same basic unit of global information, which is the position of the centroid of a cluster of points. In this fourth experiment, two sets of conditions were designed to vary the number of local information

and the number of global information. By measuring performance in both the global and local task in each condition, I tested several hypotheses regarding how VWM was occupied by both kind of information. The result of the fourth experiment supports the hypothesis best that local and global (summary) information share a unified capacity limitation in VWM.

In this dissertation, I explored the roles of global information processing, visual working memory and eye-movement in visual complex problem solving using the Traveling Salesman Problem (TSP) as a platform, producing a number of novel and theoretically interesting results. But I hope this dissertation spark people's curiosity to a new beginning and that they further explore the interaction among global information processing, visual working memory and eye-movements.

There are a few interesting open questions that naturally follow this dissertation research regarding how global representations in visual working memory can help visual problem solving. First, in our second model of Traveling Salesman Problem solving, when predicting the relationship between visual working memory capacity and eye-movement pattern, although our model predicted the trend of increasing visual working memory would generate increasing decay rate in the fixation histogram, but the prediction is shifted in term of exact visual working memory capacity. For example, to generate the same decay rate as human participants with visual working memory size around 3.5, our model's visual working memory size was set to 7. Does this suggest that the change detection task systematically underestimates visual working memory capacity, as argued by Hartshorne [74], or this suggests some underlying representation difference between our model and human? Second, in this dissertation I was using the Traveling Salesman Problem as a platform to investigate global representation in problem solving. One natural extension of this work or an open question would be how global representation is used in

other types of visual problem solving or visual tasks, such as human computer interaction, reading, driving and etc. Third, in my second model of Traveling Salesman Problem solving, I assumed that both global and local information would share a unified capacity limit in visual working memory. I also assumed that global information of different complexity, such as clusters of different sizes in this case, also share a unified capacity limit. I tested the first assumption but not the second in Experiment 4. So the assumption that global information of different sizes share a unified capacity limit in visual working memory remains untested. Fourth, in this dissertation, global information in both Traveling Salesman Problem and Multiple Objects Tracking was based on positional information and was represented by centroid location. One hypothesis is that global information is formed from similarities among local information. In this case, the similarity is location, because points in a cluster are close to each other. What about other type of global information based on similarity of other features, such as colors that are close to each other? What kind of global information may be formed from them? And how do those types of global information guide human visual problem solving in different scenarios?

I would like to end this dissertation using words from T. S. Eliot: "We shall not cease from exploration. And the end of all our exploring will be to arrive where we began and to know the place for the first time."

BIBLIOGRAPHY

1. Alvarez, G.A. and A. Oliva, *The representation of simple ensemble visual features outside the focus of attention.* Psychological Science, 2008. **19**(4): p. 392-398.

2. Baddeley, A., *Working memory.* Science, 1992. **255**(5044): p. 556-559.

3. Luck, S.J. and E.K. Vogel, *The capacity of visual working memory for features and conjunctions.* Nature, 1997. **390**(6657): p. 279-281.

4. Cowan, N., *The magical number 4 in short-term memory: a reconsideration of mental storage capacity.* Behavior Brain Science, 2001. **24**(1): p. 87-114; discussion 114-85.

5. Cowan, N., *Working memory capacity*. Essays in cognitive psychology. 2005, New York: Psychology Press. ix, 246.

6. Cowan, N., et al., *On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes.* Cognitive Psychology, 2005. **51**(1): p. 42-100.

7. Cowan, N., et al., *Theory and measurement of working memory capacity limits*, in *The psychology of learning and motivation: Advances in research and theory (Vol 49)*, B.H. Ross, Editor. 2008, Elsevier Academic Press: San Diego, CA. p. 49-104.

8. Luck, S.J. and A.R. Hollingworth, *Visual memory*. Oxford series in visual cognition ; bk. 5. 2008, Oxford ; New York: Oxford University Press. viii, 338.

9. Graham, S.M., A. Joshi, and Z. Pizlo, *The traveling salesman problem: A hierarchical model.* Memory & Cognition, 2000. **28**(7): p. 1191-1204.

10. Vickers, D., et al., *The perception of minimal structures: Performance on open and closed versions of visually presented Euclidean travelling salesperson problems.* Perception, 2003. **32**(7): p. 871-886.

11. Vickers, D., et al., *Human performance on visually presented Traveling Salesman problems.* Psychological Research/Psychologische Forschung, 2001. **65**(1): p. 34-45.

12. Pizlo, Z. and Z. Li, *Solving combinatorial problems: The 15-puzzle.* Memory & Cognition, 2005. **33**(6): p. 1069-1084.

13. de Renzi, E. and P. Nichelli, *Verbal and non-verbal short-term memory impairment following hemispheric damage.* Cortex, 1975. **11**(4): p. 341-354.

14. Calkins, M.M., *Short Studies in Memory and in Association from the Wellesley College Psychological Laboratory*. The Psychological Review, 1898. **V.**(No.5).

15. Jacobs, J., *Experiments on "Prehension"*. Mind, 1887. **os-12**(45): p. 75-79.

16. Baddeley, A., *The fractionation of working memory.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(24): p. 13468-13472.

17. Newell, A. and H.A. Simon, *Human problem solving*. 1972, Human problem solving. xiv, 920 pp. Oxford, England: Prentice-Hall Prentice-Hall Print.

18. Vogel, E.K., G.F. Woodman, and S.J. Luck, *The time course of consolidation in visual working memory.* Journal of Experimental Psychology: Human Perception and Performance, 2006. **32**(6): p. 1436-1451.

19. Shibuya, H. and C. Bundesen, *Visual selection from multielement displays: Measuring and modeling effects of exposure duration.* Journal of Experimental Psychology: Human Perception and Performance, 1988. **14**(4): p. 591-600.

20. Gegenfurtner, K.R. and G. Sperling, *Information transfer in iconic memory experiments.* Journal of Experimental Psychology: Human Perception and Performance, 1993. **19**(4): p. 845-866.

21. Atkinson, R.C. and R.M. Shiffrin, *Human memory: A proposed system and its control processes*, in *Spence, Kenneth W; Spence, Janet T. (1968).* The psychology of learning and motivation: II. xi, 249 pp. Oxford, England: Academic Press. Academic Press Print. 1968.

22. Waugh, N.C. and D.A. Norman, *Primary memory.* Psychological Review, 1965. **72**(2): p. 89-104.

23. Phillips, W.A., *On the distinction between sensory storage and short-term visual memory.* Perception & Psychophysics, 1974. **16**(2): p. 283-290.

24. Pashler, H., *Familiarity and visual change detection.* Perception & Psychophysics, 1988. **44**(4): p. 369-378.

25. Posner, M.I. and O.S.M. Marin, *Attention and performance XI*. 1985, Hillsdale, N.J.: L. Erlbaum Associates. xxiii, 675.

26. Klein, R.M., *Inhibition of return.* Trends in Cognitive Sciences, 2000. **4**(4): p. 138-147.

27. Tipper, S.P., B. Weaver, and F.L. Watson, *Inhibition of return to successively cued spatial locations: Commentary on Pratt and Abrams (1995).* Journal of Experimental Psychology: Human Perception and Performance, 1996. **22**(5): p. 1289-1293.

28. Snyder, J.J. and A. Kingstone, *Inhibition of return at multiple locations in visual search: When you see it and when you don't.* The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 2001. **54A**(4): p. 1221-1237.

29. Castel, A.D., J. Pratt, and F.I.M. Craik, *The role of spatial working memory in inhibition of return: Evidence from divided attention tasks.* Perception & Psychophysics, 2003. **65**(6): p. 970-981.

30. Alvarez, G.A. and P. Cavanagh, *The Capacity of Visual Short Term Memory Is Set Both by Visual Information Load and by Number of Objects.* Psychological Science, 2004. **15**(2): p. 106-111.

31. Awh, E., B. Barton, and E.K. Vogel, *Visual working memory represents a fixed number of items regardless of complexity.* Psychological Science, 2007. **18**(7): p. 622-628.

32. Bays, P.M. and M. Husain, *Dynamic Shifts of Limited Working Memory Resources in Human Vision.* Science, 2008. **321**(5890): p. 851-854.

33. Rouder, J.N., et al., *An assessment of fixed-capacity models of visual working memory.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(16): p. 5975-9.

34. Zhang, W. and S.J. Luck, *Discrete fixed-resolution representations in visual working memory.* Nature, 2008. **453**(7192): p. 233-5.

35. Pylyshyn, Z., *The role of location indexes in spatial perception: A sketch of the FINST spatial-index model.* Cognition, 1989. **32**(1): p. 65-97.

36. Pylyshyn, Z.W. and R.W. Storm, *Tracking multiple independent targets: Evidence for a parallel tracking mechanism.* Spatial Vision, 1988. **3**(3): p. 179-197.

37. Cavanagh, P. and G.A. Alvarez, *Tracking multiple targets with multifocal attention.* Trends in Cognitive Sciences, 2005. **9**(7): p. 349-354.

38. Scholl, B.J., *Objects and attention: The state of the art.* Cognition, 2001. **80**(1-2): p. 1-46.

39. Fougnie, D. and R. Marois, *Distinct Capacity Limits for Attention and Working Memory: Evidence From Attentive Tracking and Visual Working Memory Paradigms.* Psychological Science, 2006. **17**(6): p. 526-534.

40. Fougnie, D. and R. Marois, *Attentive tracking disrupts feature binding in visual working memory.* Visual Cognition, 2009. **17**(1-2): p. 48-66.

41. Awh, E. and J. Jonides, *Spatial working memory and spatial selective attention,* in *The attentive brain,* R. Parasuraman, Editor. 1998, The MIT Press: Cambridge, MA. p. 353-380.

42. Oh, S.-H. and M.-S. Kim, *The role of spatial working memory in visual search efficiency.* Psychonomic Bulletin & Review, 2004. **11**(2): p. 275-281.

43. Woodman, G.F. and S.J. Luck, *Visual search is slowed when visuospatial working memory is occupied.* Psychonomic Bulletin & Review, 2004. **11**(2): p. 269-274.

44. Rensink, R.A., *Visual search for change: A probe into the nature of attentional processing.* Visual Cognition, 2000. **7**(1-3): p. 345-376.

45. Rensink, R.A., *The dynamic representation of scenes.* Visual Cognition, 2000. **7**(1-3): p. 17-42.

46. Wertheimer, M., *Gestalt theory*, in *The history of psychology: Fundamental questions*. 2003, Oxford University Press: New York, NY. p. 308-323.

47. MacGregor, J.N. and T. Ormerod, *Human performance on the traveling salesman problem.* Perception & Psychophysics, 1996. **58**(4): p. 527-539.

48. Kong, X. and C.D. Schunn, *Global vs. local information processing in visual/spatial problem solving: The case of traveling salesman problem.* Cognitive Systems Research, 2007. **8**(3): p. 192-207.

49. MacGregor, J.N., E.P. Chronicle, and T.C. Ormerod, *Convex hull or crossing avoidance? Solution heuristics in the traveling salesperson problem.* Memory & Cognition, 2004. **32**(2): p. 260-270.

50. MacGregor, J.N., T.C. Ormerod, and E.P. Chronicle, *A model of human performance on the traveling salesperson problem.* Memory & Cognition, 2000. **28**(7): p. 1183-1190.

51. Rosenkrantz, D.J., et al., *An Analysis of Several Heuristics for the Traveling Salesman Problem.* SIAM Journal on Computing, 1977. **6**(3): p. 563-581.

52. Ormerod, T.C. and E.P. Chronicle, *Global perceptual processing in problem solving: The case of the traveling salesperson.* Perception & Psychophysics, 1999. **61**(6): p. 1227-1238.

53. Golden, B., et al., *Approximate Traveling Salesman Algorithms.* Operations Research, 1980. **28**(3): p. 694-711.

54. Mitchell, T.M., et al., *Predicting human brain activity associated with the meanings of nouns.* Science, 2008. **320**(5880): p. 1191-1195.

55. MacQueen, J.B., *Some methods for classification and analysis of multivariate observations.* Proc. Fifth Berkeley Symp. on Math. Statist. and Prob., 1967. **1**: p. 281-297.

56. Xu, Y. and M.M. Chun, *Visual grouping in human parietal cortex.* PNAS Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(47): p. 18766-18771.

57. de Fockert, J.W., et al., *The role of working memory in visual selective attention.* Science, 2001. **291**(5509): p. 1803-1806.

58. Woodman, G.F. and S.J. Luck, *Do the contents of visual working memory automatically influence attentional selection during visual search?* Journal of Experimental Psychology: Human Perception and Performance, 2007. **33**(2): p. 363-377.

59. Roelfsema, P.R., P.S. Khayat, and H. Spekreijse, *Subtask sequencing in the primary visual cortex.* Proc Natl Acad Sci U S A, 2003. **100**(9): p. 5467-72.

60. Sugase, Y., et al., *Global and fine information coded by single neurons in the temporal visual cortex.* Nature, 1999. **400**(6747): p. 869-873.

61. Coren, S. and P. Hoenig, *Effect of non-target stimuli upon length of voluntary saccades.* Percept Mot Skills, 1972. **34**(2): p. 499-508.

62. Findlay, J.M., *Global visual processing for saccadic eye movements.* Vision Res, 1982. **22**(8): p. 1033-45.

63. Findlay, J.M. and R. Walker, *A model of saccade generation based on parallel processing and competitive inhibition.* Behav Brain Sci, 1999. **22**(4): p. 661-74; discussion 674-721.

64. Ottes, F.P., J.A. Van Gisbergen, and J.J. Eggermont, *Metrics of saccade responses to visual double stimuli: Two different modes.* Vision Research, 1984. **24**(10): p. 1169-1179.

65. Ferguson, G.A., *Statistical analysis in psychology and education*. 5th ed. 1981, New York: McGraw-Hill. 196 p.

66. Golden, B., et al., *Approximate Traveling Salesman Algorithms.* Operations Research, 1980. **28**(3): p. 694-711.

67. Daniel, J.R., et al., *An Analysis of Several Heuristics for the Traveling Salesman Problem.* SIAM Journal on Computing, 1977. **6**(3): p. 563-581.

68. Hodgson, T.L., et al., *Eye movements and spatial working memory inParkinsons disease.* Neuropsychologia, 1999. **37**: p. 927-938.

69. Epelboim, J. and P. Suppes, *A model of eye movements and visual working memory during problem solving in geometry.* Vision Research, 2001. **41**(12): p. 1561-74.

70. Postle, B.R., et al., *The selective disruption of spatial working memory by eye movements.* Quarterly Journal of Experimental Psychology, 2006. **59**(1): p. 100-20.

71. Soto, D., G.W. Humphreys, and P. Rotshtein, *Dissociating the neural mechanisms of memory-based guidance of visual selection.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(43): p. 17186-91.

72. Brainard, D.H., *The Psychophysics Toolbox.* Spatial Vision, 1997. **10**(4): p. 433-6.
73. Pelli, D.G., *The VideoToolbox software for visual psychophysics: transforming numbers into movies.* Spatial Vision, 1997. **10**(4): p. 437-42.

74. Hartshorne, J.K., *Visual Working Memory Capacity and Proactive Interference.* PLoS ONE, 2008. **3**(7): p. e2716.